

FEATURES:
 1. 277
 273
 2056
 2056-2112



SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

Published:

— without international search report and to be republished upon receipt of that report

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

ISOLATED HUMAN TRANSPORTER PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS, AND USES THEREOF

RELATED APPLICATIONS

The present application is a Continuation-In-Part of U.S. Serial No.09/630,719, filed August 2, 2000 and Continuation-In-Part of U.S. Serial No. 09/765,344.

FIELD OF THE INVENTION

The present invention is in the field of transporter proteins that are related to the sulfate transporter subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel peptides and proteins, representing two splice forms of a novel sulfate transporter, that effect ligand transport and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

Transporters

Transporter proteins regulate many different functions of a cell, including cell proliferation, differentiation, and signaling processes, by regulating the flow of molecules such as ions and macromolecules, into and out of cells. Transporters are found in the plasma membranes of virtually every cell in eukaryotic organisms. Transporters mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of molecules and ion across cell membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, transporters, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.

Transporters are generally classified by structure and the type of mode of action. In addition, transporters are sometimes classified by the molecule type that is transported, for example, sugar transporters, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of molecule (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters: Receptor and transporter nomenclature

supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 (1997) and <http://www-biology.ucsd.edu/~msaier/transport/titlepage2.html>.

Ion channels

An important type of transporter is the ion channel. Ion channels regulate many different cell proliferation, differentiation, and signaling processes by regulating the flow of ions into and out of cells. Ion channels are found in the plasma membranes of virtually every cell in eukaryotic organisms. Ion channels mediate a variety of cellular functions including regulation of membrane potentials and absorption and secretion of ion across epithelial membranes. When present in intracellular membranes of the Golgi apparatus and endocytic vesicles, ion channels, such as chloride channels, also regulate organelle pH. For a review, see Greger, R. (1988) Annu. Rev. Physiol. 50:111-122.

Ion channels are generally classified by structure and the type of mode of action. For example, extracellular ligand gated channels (ELGs) are comprised of five polypeptide subunits, with each subunit having 4 membrane spanning domains, and are activated by the binding of an extracellular ligand to the channel. In addition, channels are sometimes classified by the ion type that is transported, for example, chlorine channels, potassium channels, etc. There may be many classes of channels for transporting a single type of ion (a detailed review of channel types can be found at Alexander, S.P.H. and J.A. Peters (1997). Receptor and ion channel nomenclature supplement. Trends Pharmacol. Sci., Elsevier, pp. 65-68 and <http://www-biology.ucsd.edu/~msaier/transport/toc.html>.

There are many types of ion channels based on structure. For example, many ion channels fall within one of the following groups: extracellular ligand-gated channels (ELG), intracellular ligand-gated channels (ILG), inward rectifying channels (INR), intercellular (gap junction) channels, and voltage gated channels (VIC). There are additionally recognized other channel families based on ion-type transported, cellular location and drug sensitivity. Detailed information on each of these, their activity, ligand type, ion type, disease association, drugability, and other information pertinent to the present invention, is well known in the art.

Anion transport proteins

The present invention provides two splice forms of a novel human anion transport protein that shows a particularly high degree of similarity to sulfate transporters. The alternative splice forms are herein referred to as splice forms 1 and 2. Splice form 1 has been previously disclosed by applicant in U.S. application 09/630,719, filed August 2, 2000.

Anion transport proteins in mammalian cells participate in a wide variety of cell and intracellular organelle functions, including regulation of electrical activity, pH, volume, and the transport of osmolites and metabolites. These proteins also have essential physiological roles in the control of immunological responses, cell migration, cell proliferation, and differentiation. Several classes of anion transporters have been characterized with varying molecular structures and mechanisms for mediating anion flux. One of the most prominent anion transporter super-families is the multiple membrane-spanning permeases, which include Na⁺- or H⁺-dependent anion coanion transporters (symporters), anion/anion exchangers (antiporters), and cation-independent anion uniporters. This super-family is also referred to as the Carrier-type anion transporters <<http://www-biology.ucsd.edu/~msaier/transport/titlepage2.html>>. Typically, proteins within this group contain 8-14 hydrophobic alpha-helical peptide segments that allow the protein to reside in the membrane bilayer. These helices also establish the pathway for ion translocation. Both broad-substrate and substrate-specific anion transporters are known: the former type enables multiple anion species (chloride, iodide, sulfate, bromide, etc.) to permeate the bounding membrane, while anion transporters in the latter class restrict ion movement to one chemical species.

Anion transporter genes and gene products are potential causative agents of disease and disease phenotypes may be actuated both by alterations in gene transcription and by mutations in the protein sequence. For example, the down-regulated in adenoma (DRA) gene was originally identified as a gene that was down-regulated in colon tumors. It encodes a protein with anion transporter function that is expressed in the intestinal tract (duodenum, ileum, cecum, distal colon), but not in the esophagus or stomach (Antalis T.M., Reeder J.A., Gotley D.C. *et al.*, *Clin Cancer Res* (1998) Aug;4(8):1857-63; Byeon M.K., Westerman M.A., Maroulakou I.G. *et al.*, *Oncogene* (1996) Jan 18;12(2):387-96). A second illustration of the biomedical significance of anion transporters is found with patients presenting severe hypothyroidism caused by a congenital lack of iodide transport. These individuals do not accumulate-iodide in their thyroids. A single amino acid substitution in the thyroid Na⁺/I⁻ symporter, where proline replaced threonine at position 354, has been identified as the cause of this condition in two independent patients (Levy O., Ginter C.S., De la Vieja A. *et al.*, *FEBS Lett* (1998) Jun 5;429(1):36-40). Equally compelling are two well-documented autosomal recessive disorders Pendred syndrome and Diastrophic dysplasia (DTD). Pendred syndrome is the most common form of syndromic deafness and characterized by congenital sensorineural hearing loss and goitre. This disorder has been mapped to chromosome 7 and the gene product causing Pendred syndrome (PDS) has been

identified as a anion transporter for iodide and chloride (Scott D.A., Wang R., Kreman T.M. *et al.*, *Nat Genet* (1999)Apr;21(4):440-3). DTD is a well-characterized osteochondrodysplasia with clinical features including dwarfism, spinal deformation, and specific joint abnormalities. The disease occurs in most populations. The gene has been mapped to distal chromosome 5q and it encodes a sulfate anion transporter (Hastbacka J., de la Chapelle A., Mahtani M.M. *et al.*, *Cell* 1994 Sep 23;78(6):1073-87).

Issued US Patents that demonstrate the utility for this group of protein/DNA molecules include, but are not limited to, 6,054,558 "Compositions and methods for the treatment and diagnosis of cardiovascular disease using rchd534 as a target"; 6,048,709 "Compositions and methods for the treatment and diagnosis of cardiovascular disease"; 6,046,030 "Human LIG-1 homolog (HLIG-1)"; 6,025,160 "Polynucleotide and polypeptide sequences encoding rat mdr1b2 and screening methods thereof"; 6,013,672 "Agonists of metabotropic glutamate receptors and uses thereof"; 6,008,015 "Glycine transporter"; 5,989,825 "Excitatory amino acid transporter gene and uses"; and 5,928,926 "Isolation and cloning of the human ARSA-I gene and uses thereof".

Transporter proteins, particularly members of the sulfate transporter subfamily, are a major target for drug action and development, particularly members that are expressed in the tissue types noted in Figure 1 (e.g. neoplastic cells)). Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown transport proteins. The present invention advances the state of the art by providing a previously unidentified human transport protein.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of amino acid sequences of human transporter peptides and proteins that are related to the sulfate transporter subfamily, as well as allelic variants and other mammalian orthologs thereof. Specifically, the present invention provides two splice forms of a novel human sulfate transporter protein. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents that modulate transporter activity in cells and tissues that express the transporter. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal

muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas.

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequences of cDNA molecules that encode splice forms 1 and 2 of the transporter protein of the present invention (splice form 1 = SEQ ID NO:1, splice form 2 = SEQ ID NO:4). In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of the inventions based on these molecular sequences. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas.

FIGURE 2 provides the predicted amino acid sequences of splice forms 1 and 2 of the transporter of the present invention (splice form 1 = SEQ ID NO:2, splice form 2 = SEQ ID NO:5). In addition, structure and functional information such as protein family, function, and modification sites is provided where available, allowing one to readily determine specific uses of the inventions based on these molecular sequences.

FIGURE 3 provides a genomic sequence SEQ ID NO:3) that spans the gene encoding splice forms 1 and 2 of the transporter protein of the present invention. In addition, structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of the inventions based on this molecular sequence. As illustrated in Figure 3, identified SNP variations include g30344a, a31170g, c16256t, a13376t, t12210c, g12072c, g11922t, -11903a, c10009g, c4519t, a4181g, a20952c, t20987c, g21620a, t21795c, a22753t, g22945a, g23032a, g23738a, t23952g, a24123g, c24527-, c24691t, g25015a, and g25191t. Figure 3 also provides structural information for splice form 2, derived from the Genewise computer program.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a transporter protein or part of a transporter protein and are related to the sulfate transporter subfamily. Specifically, the present invention provides two splice forms of a novel human sulfate transporter. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human transporter peptides and proteins that are related to the sulfate transporter subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these transporter peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the transporter of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known transporter proteins of the sulfate transporter subfamily and the expression pattern observed. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas.. The art has clearly established the commercial importance of members of this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known sulfate transporter family or subfamily of transporter proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the transporter family of proteins and are related to the sulfate transporter subfamily (protein sequences are provided in Figure 2, cDNA sequences are provided in Figures 1 and genomic sequences are provided in Figure 3). Specifically, the present invention provides two splice forms of a novel human sulfate transporter. The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the transporter peptides of the present invention, transporter peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprising the amino acid sequences of the transporter peptides disclosed in the Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical

precursors or other chemicals" includes preparations of the transporter peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated transporter peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. For example, a nucleic acid molecule encoding the transporter peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NOS:2 and 5), for example, proteins encoded by the cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1 and 4) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NOS:2 and 5), for example, proteins encoded by the cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1 and 4) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NOS:2 and 5), for example, proteins encoded by the cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NOS:1 and 4) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid

sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the transporter peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The transporter peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a transporter peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the transporter peptide. "Operatively linked" indicates that the transporter peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the transporter peptide.

In some uses, the fusion protein does not affect the activity of the transporter peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant transporter peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A transporter peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the transporter peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring mature

forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the transporter peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of a reference sequence is aligned for comparison purposes. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm

which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al.*, *Nucleic Acids Res.* 12(1):387 (1984)) (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Myers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, *et al.* (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.* (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the transporter peptides of the present invention as well as being encoded by the same genetic locus as the transporter peptide provided herein. RH panel mapping shows the gene encoding the transporter proteins of the present invention is found on chromosome 17 near markers SHGC-56719 and SHGC-58932 (LOD=15.9 and 15.68).

Allelic variants of a transporter peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by the same genetic locus as the transporter peptide provided herein. Genetic locus can readily be determined based on the genomic information

provided in Figure 3, such as the genomic sequence mapped to the reference human. RH panel mapping shows the gene encoding the transporter proteins of the present invention is found on chromosome 17 near markers SHGC-56719 and SHGC-58932 (LOD=15.9 and 15.68). As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been found in the gene encoding the transporter proteins of the present invention. The following variations were seen: g30344a, a31170g, c16256t, a13376t, t12210c, g12072c, g11922t, -11903a, c10009g, c4519t, a4181g, a20952c, t20987c, g21620a, t21795c, a22753t, g22945a, g23032a, g23738a, t23952g, a24123g, c24527-, c24691t, g25015a, and g25191t.

Paralogs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a transporter peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the transporter peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a transporter peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the transporter peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the transporter peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a transporter peptide by another amino acid of like

characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

Variant transporter peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to transport ligand, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis and can be used to identify critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as transporter activity or in assays such as an *in vitro* proliferative activity. Sites that are critical for binding partner/substrate binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol.* 224:899-904 (1992); de Vos *et al.* *Science* 255:306-312 (1992)).

The present invention further provides fragments of the transporter peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a transporter peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the transporter peptide or could be chosen for the

ability to perform a function, e.g. bind a substrate or act as an immunogen. Particularly important fragments are biologically active fragments, peptides that are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the transporter peptide, e.g., active site, a transmembrane domain or a substrate-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in transporter peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art (some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol.* 182: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci.* 663:48-62 (1992)).

Accordingly, the transporter peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature transporter peptide is fused with another compound, such as a compound to increase the half-life of the transporter peptide (for example, polyethylene glycol), or in which the additional amino acids are fused to the mature transporter peptide, such as a leader or secretory sequence or a sequence for purification of the mature transporter peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or ligand) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein or ligand (such as, for example, in a transporter-effector protein interaction or transporter-ligand interaction), the protein can be used to identify the binding partner/ligand so as to develop a system to identify inhibitors of the binding interaction. Any or all of these uses are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, transporters isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the transporter. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas.

Additionally, BLAST hits to ESTs derived from tissue specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen. A large percentage of pharmaceutical agents are being developed that modulate the activity of transporter proteins, particularly members of the sulfate transporter subfamily (see Background of the Invention). The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemia B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. Such uses can readily be determined using the information provided herein, that known in the art and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to transporters that are related to members of the sulfate transporter subfamily. Such assays involve any of the known transporter functions or activities or properties useful for diagnosis and treatment of transporter-related conditions that are specific for the subfamily of transporters that the one of the present invention belongs to, particularly in cells and tissues that express the transporter. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas. Additionally, BLAST hits to ESTs derived from tissue specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen. The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems ((Hodgson, Bio/technology, 1992, Sept 10(9);973-80). Cell-based systems can be native, i.e., cells that normally express the transporter, as a biopsy or expanded in cell culture. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated

tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the transporter protein.

The polypeptides can be used to identify compounds that modulate transporter activity of the protein in its natural state or an altered form that causes a specific disease or pathology associated with the transporter. Both the transporters of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the transporter. These compounds can be further screened against a functional transporter to determine the effect of the compound on the transporter activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the transporter to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the transporter protein and a molecule that normally interacts with the transporter protein, e.g. a substrate or a component of the signal pathway that the transporter protein normally interacts (for example, another transporter). Such assays typically include the steps of combining the transporter protein with a candidate compound under conditions that allow the transporter protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the transporter protein and the target, such as any of the associated effects of signal transduction such as changes in membrane potential, protein phosphorylation, cAMP turnover, and adenylate cyclase activation, etc.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature* 354:82-84 (1991); Houghten *et al.*, *Nature* 354:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell* 72:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, F(ab')₂, Fab expression library

fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant transporters or appropriate fragments containing mutations that affect transporter function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) transporter activity. The assays typically involve an assay of events in the signal transduction pathway that indicate transporter activity. Thus, the transport of a ligand, change in cell membrane potential, activation of a protein, a change in the expression of genes that are up- or down-regulated in response to the transporter protein dependent signal cascade can be assayed.

Any of the biological or biochemical functions mediated by the transporter can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the transporter can be assayed. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas. Additionally, BLAST hits to ESTs derived from tissue specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen.

Binding and/or activating compounds can also be screened by using chimeric transporter proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a ligand-binding region can be used that interacts with a different ligand than that which is recognized by the native transporter. Accordingly, a different set of signal transduction components is available as an end-

point assay for activation. This allows for assays to be performed in other than the specific host cell from which the transporter is derived.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the transporter (e.g. binding partners and/or ligands). Thus, a compound is exposed to a transporter polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide. Soluble transporter polypeptide is also added to the mixture. If the test compound interacts with the soluble transporter polypeptide, it decreases the amount of complex formed or activity from the transporter target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the transporter. Thus, the soluble polypeptide that competes with the target transporter region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the transporter protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ³⁵S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of transporter-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a transporter-binding protein and a candidate compound are incubated in the transporter protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the

GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the transporter protein target molecule, or which are reactive with transporter protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the transporters of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of transporter protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the transporter pathway, by treating cells or tissues that express the transporter. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. These methods of treatment include the steps of administering a modulator of transporter activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the transporter proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos *et al.* (1993) *Cell* 72:223-232; Madura *et al.* (1993) *J. Biol. Chem.* 268:12046-12054; Bartel *et al.* (1993) *Biotechniques* 14:920-924; Iwabuchi *et al.* (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the transporter and are involved in transporter activity. Such transporter-binding proteins are also likely to be involved in the propagation of signals by the transporter proteins or transporter targets as, for example, downstream elements of a transporter-mediated signaling pathway. Alternatively, such transporter-binding proteins are likely to be transporter inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a transporter protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an

unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a transporter-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the transporter protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a transporter-modulating agent, an antisense transporter nucleic acid molecule, a transporter-specific antibody, or a transporter-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The transporter proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. The method involves contacting a biological sample with a compound capable of interacting with the transporter protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered transporter activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from

standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the transporter protein in which one or more of the transporter functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and transporter activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. Accordingly, methods for treatment include the use of the transporter protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or F(ab')₂, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, *Antibodies*, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the transporter proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or transporter/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas. Additionally, BLAST hits to ESTs derived from tissue specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution, developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Independent lines of evidence show expression of splice

form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemic B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the transporter peptide to a binding partner such as a ligand or protein binding partner. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount

of protein in the sample; means for comparing the amount of protein in the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a transporter peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the transporter peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences that naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NOS:1 and 4, cDNA/transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NOS:2 and 5. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NOS:1 and 4, cDNA/transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NOS:2 and 5. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NOS:1 and 4, cDNA/transcript sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NOS:2 and 5. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprise several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, humans genomic sequence (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the mature form

has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the transporter peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the transporter proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression and in

developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could be at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. RH panel mapping shows the gene encoding the transporter proteins of the present invention is found on chromosome 17 near markers SHGC-56719 and SHGC-58932 (LOD=15.9 and 15.68).

Figure 3 provides information on SNPs that have been found in the gene encoding the transporter proteins of the present invention. The following variations were seen: g30344a, a31170g, c16256t, a13376t, t12210c, g12072c, g11922t, -11903a, c10009g, c4519t, a4181g, a20952c, t20987c, g21620a, t21795c, a22753t, g22945a, g23032a, g23738a, t23952g, a24123g, c24527-, c24691t, g25015a, and g25191t.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more

homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, identified SNP variations include g30344a, a31170g, c16256t, a13376t, t12210c, g12072c, g11922t, -11903a, c10009g, c4519t, a4181g, a20952c, t20987c, g21620a, t21795c, a22753t, g22945a, g23032a, g23738a, t23952g, a24123g, c24527-, c24691t, g25015a, and g25191t.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. RH panel

mapping shows the gene encoding the transporter proteins of the present invention is found on chromosome 17 near markers SHGC-56719 and SHGC-58932 (LOD=15.9 and 15.68).

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas. Additionally, BLAST hits to ESTs derived from tissue specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen.

Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in transporter protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA include Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a transporter protein, such as by measuring a level of a transporter-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a transporter gene

has been mutated. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas. Additionally, BLAST hits to ESTs derived from tissue specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate transporter nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the transporter gene, particularly biological and pathological processes that are mediated by the transporter in cells and tissues that express it. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemia B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas. The method typically includes assaying the ability of the compound to modulate the expression of the transporter nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired transporter nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the transporter nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for transporter nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the transporter protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of transporter gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of transporter mRNA in the presence of the candidate compound is compared to the level of expression of transporter mRNA in the absence of the candidate compound. The candidate

compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate transporter nucleic acid expression in cells and tissues that express the transporter. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas. Additionally, BLAST hits to ESTs derived from tissue specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) of nucleic acid expression.

Alternatively, a modulator for transporter nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the transporter nucleic acid expression in the cells and tissues that express the protein. Independent lines of evidence show expression of splice form 1 in fetal tissues such as brain and kidney; differentiated tissues such as brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas; and neoplastic tissues such as germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendrogliomas, chronic lymphocytic leukemia B-cells, glioblastomas, and well-differentiated endometrial adenocarcinomas.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the transporter gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a

physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in transporter nucleic acid expression, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in transporter genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally occurring genetic mutations in the transporter gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the transporter gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a transporter protein.

Individuals carrying mutations in the transporter gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter proteins of the present invention. The following variations were seen: g30344a, a31170g, c16256t, a13376t, t12210c, g12072c, g11922t, -11903a, c10009g, c4519t, a4181g, a20952c, t20987c, g21620a, t21795c, a22753t, g22945a, g23032a, g23738a, t23952g, a24123g, c24527-, c24691t, g25015a, and g25191t. RH panel mapping shows the gene encoding the transporter proteins of the present invention is found on chromosome 17 near markers SHGC-56719 and SHGC-58932 (LOD=15.9 and 15.68). Genomic DNA can be analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science* 241:1077-1080 (1988); and Nakazawa *et al.*, *PNAS* 91:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res.* 23:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample,

contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a transporter gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant transporter gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques* 19:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr.* 36:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol.* 38:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science* 230:1242 (1985)); Cotton *et al.*, *PNAS* 85:4397 (1988); Saleeba *et al.*, *Meth. Enzymol.* 217:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS* 86:2766 (1989); Cotton *et al.*, *Mutat. Res.* 285:125-144 (1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* 9:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* 313:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the

individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the transporter gene in an individual in order to select an appropriate compound or dosage regimen for treatment. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter proteins of the present invention. The following variations were seen: g30344a, a31170g, c16256t, a13376t, t12210c, g12072c, g11922t, -11903a, c10009g, c4519t, a4181g, a20952c, t20987c, g21620a, t21795c, a22753t, g22945a, g23032a, g23738a, t23952g, a24123g, c24527-, c24691t, g25015a, and g25191t.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control transporter gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of transporter protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into transporter protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of transporter nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired transporter nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the transporter protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in transporter gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired transporter protein to treat the individual.

The invention also encompasses kits for detecting the presence of a transporter nucleic acid in a biological sample. Screening of tissue specific cDNA libraries for cDNA retrieval showed expression of splice form 1 in fetal brain, brain, pituitary gland, heart, leukocytes, kidney, liver, thyroid, lung, placenta, skeletal muscle, fetal kidney, small intestine, prostate, testis, adrenal gland, bone marrow, and pancreas. Additionally, BLAST hits to ESTs derived from tissue

specific libraries shows expression of splice form 1 in germ cell tumors, lung carcinoid tissue, brain anaplastic oligodendroglioma, chronic lymphocytic leukemia B-cells, infant brain (73 days post natal), placenta, glioblastoma, well-differentiated endometrial adenocarcinoma, and fetal liver and spleen. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting transporter nucleic acid in a biological sample; means for determining the amount of transporter nucleic acid in the sample; and means for comparing the amount of transporter nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect transporter protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1, 3, and 4).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee *et al.*, PCT application W095/11995 (Chee *et al.*), Lockhart, D. J. *et al.* (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. *et al.* (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown *et al.*, US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides that cover the full length sequence; or unique oligonucleotides selected from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler *et al.*) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative

abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large-scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the transporter proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the transporter gene of the present invention. Figure 3 provides information on SNPs that have been found in the gene encoding the transporter proteins of the present invention. The following variations were seen: g30344a, a31170g, c16256t, a13376t, t12210c, g12072c, g11922t, -11903a, c10009g, c4519t, a4181g, a20952c, t20987c, g21620a, t21795c, a22753t, g22945a, g23032a, g23738a, t23952g, a24123g, c24527-, c24691t, g25015a, and g25191t.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques*, Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al.*, *Techniques in Immunocytochemistry*, Academic Press, Orlando, FL Vol. 1 (1982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology*, Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified transporter gene of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ , the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses,

papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, e.g. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd. ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterotrinsporter. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene* 67:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann

et al., *Gene* 69:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology* 185:60-89 (1990)).

Recombinant protein expression can be maximized in host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res.* 20:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J.* 6:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell* 30:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene* 54:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol.* 3:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology* 170:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature* 329:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J.* 6:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters

described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al. (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).*

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell-free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as transporters, appropriate secretion signals are incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with transporters, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a transporter protein or peptide that can be further purified to produce desired amounts of transporter protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the transporter protein or transporter protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native transporter protein is useful for assaying compounds that stimulate or inhibit transporter protein function.

Host cells are also useful for identifying transporter protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant transporter protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native transporter protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse,

in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA that is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a transporter protein and identifying and evaluating modulators of transporter protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the transporter protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the transporter protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al.* *PNAS* 89:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al.* *Science* 251:1351-1355 (1991)). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein

is required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature* 385:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, transporter protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* transporter protein function, including ligand interaction, the effect of specific mutant transporter proteins on transporter protein function and ligand interaction, and the effect of chimeric transporter proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more transporter protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:

- (a) an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;
- (b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4;
- (c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4; and
- (d) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids.

2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:

- (a) an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;
- (b) an amino acid sequence of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4;
- (c) an amino acid sequence of an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4; and
- (d) a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids.

3. An isolated antibody that selectively binds to a peptide of claim 2.
4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;
 - (b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4;
 - (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4;
 - (d) a nucleotide sequence that encodes a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids; and
 - (e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).
5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence that encodes an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5;
 - (b) a nucleotide sequence that encodes an allelic variant of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4;
 - (c) a nucleotide sequence that encodes an ortholog of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4;

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.
7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.
8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.
9. A host cell containing the vector of claim 8.
10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.
12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.
13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.

14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.
15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.
16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.
17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.
18. A method for treating a disease or condition mediated by a human transporter protein, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.
19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human transporter peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence selected from the group consisting of SEQ ID NOS:2 and 5.

22. An isolated nucleic acid molecule encoding a human transporter peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4.

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule selected from the group consisting of SEQ ID NOS:1, 3, and 4.

SPLICE FORM 1:

```

1 AACAGCACGA GGGCGGACCC AGCTGTGGCG ACGCCAGGAG ACCCCAAGCT
51 GCATCGCCGA GTGGAAGCAA CTAGAACTCC AGGGCTGTGA AAGCCACAGG
101 TGGGGGGCTGA GCGAGGCGTG GCCTCAGGAG CGGAGGACCC CCCACTCTC
151 CCTCGAGCGC CGCAGTCCAC CGTAGCGGGT GGAGCCCGCC TTGGTGCGCA
201 GTTGGAAAAC CTCGGAGCCC CGCTGGATCT CCTGGCTGCC ACCCGCACCC
251 CCCGCCAGCC TACGCCCCAC CGTAGAGATG CCTTCTTCGG TGACGGCGCT
301 GGGTCAGGCC AGGTCTCTG GCGCCGGGAT GCGCCCGAGC GCCTGCTGCT
351 GCTCCCTGTC GGCCCTGCAG AGGAGGCTGC CCATCTGGC GTGGCTGCCC
401 AGCTACTCCC TGCAGTGGCT GAAGATGGAT TTCGTGCGCC GCCTCTCAGT
451 TGGCCTCACT GCCATTCCCC AGGCGCTGGC CTATGCTGAA GTGGCTGGAC
501 TCCCGCCCCA GTATGGCCTC TACTCTGCCT TCATGGGCTG CTTGCTGTAT
551 TTCTTCTTGG GCACCTCCCG GGATGTGACT CTGGGCCCCA CCGCATTAT
601 GTCCCTCCTG GTCTCCTTCT ACACCTTCCA TGAGCCCGCC TACGCTGTGC
651 TGCTGGCCTT CCTGTCCGGC TGCATCCAGC TGGCCATGGG GGTCTGCGT
701 TTGGGGTTCC TGCTGGACTT CATTTCTTAC CCCGTCATTA AAGGCTTCAC
751 CTCTGCTGCT GCCGTCACCA TCGGCTTTGG ACAGATCAAG AACCTGCTGG
801 GACTACAGAA CATCCCCAGG CCGTTCTTCC TGCAGGTGTA CCACACCTTC
851 CTCAGGATTG CAGAGACCAG GGTAGGTGAC GCGCTCCTGG GGCTGGTCTG
901 CATGTGCTG CTGCTGGTGC TGAAGCTGAT GCGGGACCAC GTGCCTCCCG
951 TCCACCCCGA GATGCCCCCT GGTGTGCGGC TCAGCCGTGG GCTGGTCTGG
1001 GCTGCCACGA CAGCTCGCAA CGCCCTGGTG GTCTCCTTCG CAGCCCTGGT
1051 TGCTTACTCC TTCGAGGTGA CTGGATACCA GCCTTTCATC CTAACAGGGG
1101 AGACAGCTGA GGGGCTCCCT CCAGTCCGGA TCCCGCCCTT CTCAGTGACC
1151 ACAGCCAACG GGACGATCTC CTTACCGAG ATGGTGCAGG ACATGGGAGC
1201 CGGGCTGGCC GTGGTGCCCC TGATGGGCTT CCTGGAGAGC ATTGCGGTGG
1251 CCAAAGCCTT CGCATCTCAG AATAATTACC GCATCGATGC CAACCAGGAG
1301 CTGCTGGCCA TCGGTCTCAC CAACATGTTG GGCTCCCTCG TCTCCTCCTA
1351 CCCGTCACA GGCAGCTTTG GACGGACAGC CGTGAACGCT CAGTCGGGGG
1401 TGTGACCCCC GCGGGGGGGC CTGGTGACGG GAGTGCTGGT GCTGCTGTCT
1451 CTGGACTACC TGACCTCACT GTTCTACTAC ATCCCCAAGT CTGCCCTGGC
1501 TGCCCTCATC ATCATGGCCG TGGCCCGCT GTTCGACACC AAGATCTTCA
1551 GGACGCTCTG GCGTGTTAAG AGGCTGGACC TGCTGCCCT GTGCGTGACC
1601 TTCCTGCTGT GCTTCTGGGA GGTGCAGTAC GGCATCCTGG CCGGGGCCCT
1651 GGTGTCTCTG CTCATGCTCC TGCACCTGTC AGCCAGGCCT GAGACCAAGG
1701 TGTGAGAGGG GCCGTTCTG GTCTGCAGC CGGCCAGCGG CCTGTCTTTC
1751 CCTGCCATGG AGGCTCTGCG GGAGGAGATC CTAAGCCGGG CCCTGGAAGT
1801 GTCCCGGCCA CGCTGCCTGG TCCTGGAGTG CACCCATGTC TGCAGCATCG
1851 ACTACACTGT GGTGCTGGGA CTCGGCGAGC TCCTCCAGGA CTCCAGAAG
1901 CAGGGCGTCG CCCTGGCCTT TGTGGGCTG CAGGTCCCG TTCTCCGTGT
1951 CCTGCTGTCC GCTGACCTGA AGGGTTCCA GTACTTCTCT ACCCTGGAAG
2001 AAGCAGAGAA GCACCTGAGG CAGGAGCCAG GGACCCAGCC CTACAACATC
2051 AGAGAAGACT CCATTCTGGA CCAAAGGTT GCCCTGCTCA AGGCATAATG
2101 GGGCCACCCG TGGGCATCCA CAGTTTGAG GGTGTTCCGG AAGGTTCTTG
2151 TCACTGTGAT TGGATGCTGG ATGCCGCTG ATAGACATG TGGCCTGGCT
2201 GAGAAACCCC TGAGCAGGTA ACCCAGGGAA GAGAAGGAAG CCAGGCCTGG
2251 AGGTCCACGG CAGTGGGAGT GGGGCTCACT GGCTTCTGT GGGATGACTG
2301 GAAATGACC TCGCTGCTGT TCCCTGGCAT GACCTCTTT GGAAGAGTGG
2351 TTTGGAGAGA GCCTTCTAGA ATGACAGACT GTGCGAGGAA GCAGGGGCAG
2401 GGGTTTCCAG CCCGGGCTGT GCGAGGCATC CTGGGGCTGG CAGCACCTTC
2451 CCGGCTCACC AGTGCCACCT GCGGGGGAGG GACGGGGCAG GCAGGAGTCT
2501 GGGAGGCGGG TCCGCTCCTC TTGTCTGCGG CATCTGTGCT CTCCGAGAGA
2551 AAACCAAGGT GTGTCAAATG ACGTCAAGTC TCTATTAAA AATAATTTTG
2601 TGTTTTCTAA ATGGAAAAAG TGATAGCTTT GGTGATTTT TAAAAGTCAT
2651 AAATGCTTAT TGTAAAAAAT ACAGGAAACC ACCCTCACC CTGTCCACTT
2701 GGGTGATCAT TCCAGACCCC TCCCCAAACA TGCATATGTA CCTGTCCGTC
2751 AGTGTGTGGA TGTATGTTTA CAGTTCTACA TAAATGGGAT CATTTTATAC
2801 ATGGTGCTCT GGAACCCACA TTTTTCATGC AGTCATTTC AGTGAATTAT
2851 TTATTGTGAT AATAAATAGC ATTAGAATAC AAGAAAAAAA AAAAAAAA
2901 AAAAAAAA AAAAAAAA (SEQ ID NO:1)

```

FEATURES:

5' UTR: 1-277

Start: 278

Stop: 2096

3' UTR: 2096-2919

SPLICE FORM 2:

```

1   GCTGTGGCGA CGCCAGGAGA CCCCAGCTG CATCGCCGAG TGAAGCAAC
51  TAGAACTCCA GGGCTGTGAA AGCCACAGGT GGGGGCTGAG CGAGGCGTGG
101 CCTCAGGAGC GGAGGACCCC CCCACTCTCC CTCGAGCGCC GCAGTCCACC
151 GTAGCGGGTG GAGCCCGCCT TGGTGCAGAG TTGGAAGAAC TCGGAGCCCC
201 GCTGGATCTC CTGGCTGCCA CCCGCACCCC CCGCCAGCCT ACGCCCCACC
251 GTAGAGATGC CTTCTTCGGT GACGGCGCTG GGTGAGGCCA GGTCTCTGG
301 CCCCAGGATG GCCCCGAGCG CCTGTGCTG CTCCCCTGCG GCCCTGCAGA
351 GGAGGCTGCC CATCTGGCG TGGCTGCCCC GCTACTCCCT GCAGTGGCTG
401 AAGATGGATT TCGTCGCCGG CCTCTCAGTT GGCCTCACTG CCATTCCCCA
451 GGCCTGGGCC TATGCTGAAG TGGCTGGACT CCCGCCCCAG TATGGCCTCT
501 ACTCTGCCCT CATGGGCTGC TTCGTGTATT TCTTCCTGGG CACCTCCCGG
551 GATGTGACTC TGGGCCCCAC CGCCATTATG TCCCTCCTGG TCTCCTTCTA
601 CACCTTCCAT GAGCCCGCCT ACGCTGTGCT GCTGGCCTTC CTGTCCGGCT
651 GCATCCAGCT GGCCATGGGG GTCCTGCGTT TGGCTCACAT CTCCCCTCAT
701 CACTTGGGAC TGGGTGGAGC CGGACCAGC TCGATGTCCC CTCTTGGCTG
751 GCCAGGGTTC CTGCTGGACT TCATTTCCTA CCCCCTCATT AAAGGCTTCA
801 CCTCTGCTGC TGCCGTCACC ATCGGCTTTG GACAGATCAA GAACCTGCTG
851 GGACTACAGA ACATCCCCAG GCCGTCTTTC CTGCAGGTGT ACCACACCTT
901 CCTCAGGATT GCAGAGACCA GGGTAGGTGA CGCCGTCTCT GGGCTGGTCT
951 GCATGTGCTG GCTGTGGTG CTGAAGCTGA TCGGGGACCA CGTGCCTCCC
1001 GTCCACCCCG AGATGCCCCC TGGTGTGCGG CTCAGCCGTG GGCTGGTCTG
1051 GGCTGCCACG ACAGCTCGCA ACGCCCTGGT GGTCTCCTTC GCAGCCCTGG
1101 TTGCGTACTC CTTTCGAGTG ACTGGATACC AGCCTTTCAT CCTAACAGGG
1151 GAGACAGCTG AGGGGCTCCC TCCAGTCCGG ATCCCGCCCT TCTCAGTGAC
1201 CACAGCCAAC GGGACGATCT CCTTCACCGA GATGGTGAG GACATGGGAG
1251 CCGGGCTGGC CGTGGTGCCC CTGATGGGCC TCCTGGAGAG CATTGCGGTG
1301 GCCAAGCCT TCGCATCTCA GAATAATTAC CGCATCGATG CCAACCAGGA
1351 GCTGTGGGCC ATCGGTCTCA CCAACATGTT GGGCTCCCTC GTCTCCTCCT
1401 ACCCGGTCAC AGGCAGCTTT GGACGGACAG CCGTGAACGC TCAGTCGGGG
1451 GTGTGCACCC CGGCGGGGGG CCTGGTGACG GGAGTGCTGG TGCTGCTGTC
1501 TCTGGACTAC CTGACCTCAC TGTTCTACTA CATCCCCAAG TCTGCCCTGG
1551 CTGCCGTGAT CATCATGGCC GTGGCCCGCG GTTTCGACAC CAAGATCTTC
1601 AGGACGCTCT GCGGTGTTAA GAGGCTGGAC CTGCTGCCCC TGTGGGTGAC
1651 CTTCTGCTG TGCTTCTGGG AGGTGCAGTA CGGCATCCTG GCCGGGGCCC
1701 TGGTGTCTCT GCTCATGCTC CTGCACTCTG CAGCCAGGCC TGAGACCAAG
1751 GTGTGAGAGG GGCCGTTCTT GGTCTGCGAG CCGGCCAGCG GCCTGTCTCT
1801 CCCTGCCATG GAGGCTCTGC GGGAGGAGAT CCTAAGCCGG GCCCTGGAAG
1851 TGTCCCGGCC ACGCTGCCTG GTCCTGGAGT GCACCCATGT CTGCAGCATC
1901 GACTACACTG TGGTGTGGG ACTCGGCGAG CTCTCCAGG ACTTCCAGAA
1951 GCAGGGCGTC GCCCTGGCCT TTGTGGGCCT GCAGGTCCCC GTTCTCCGTG
2001 TCCTGCTGTC CGCTGACCTG AAGGGGTTC AGTACTTCTC TACCCTGGAA
2051 GAAGCAGAGA AGCACCTGAG GCAGGAGCCA GGGACCCAGC CCTACAACAT
2101 CAGAGAAGAC TCCATTCTGG ACCAAAAGGT TGGCCTGCTC AAGGCATAAT
2151 GGGGCCACCC GTGGGCATCC ACAGTTTGCA GGGTGTTCGG GAAGGTTCTT
2201 GTCATGTGTA TTGGATGCTG GATGCCGCTT GATAGACATG CTGGCCTGGC
2251 TGAGAAACCC CTGAGCAGGT AACCAGGGA AGAGAAGGAA GCCAGGCCTG
2301 GAGTCCACG GCAGTGGGAG TGGGGCTCAC TGGCTTCCTG TGGGATGACT
2351 GGAAATGAC CTCGCTGCTG TTCCCTGGCA TGACCCTCTT TGGAAGAGTG
2401 GTTTGGAGAG AGCCTTCTAG AATGACAGAC TGTGCGAGGA AGCAGGGGCA
2451 GGGGTTTCCA GCCCAGGCTG TGCGAGGCAT CCTGGGGCTG GCAGCACCTT
2501 CCCGCTCAC CAGTGCCACC TGCGGGGGAG GGACGGGGCA GGCAGGAGTC
2551 TGGGAGGCGG GTCCGCTCCT CTTGTCTGCG GCATCTGTGC TCTCCGAGAG
2601 AAAACCAAGG TGTGTCAAAT GACGTCAAGT CTCTATTATA AAATAATTTT
2651 GTGTTTCTA AATGGAAAAA GTGATAGCTT TGGTGATTTT GTAAAAGTCA
2701 TAAATGCTTA TTGTAAAAAA TACAGGAAAC CACCCCTCAC CCTGTCCACT
2751 TGGGTGATCA TTCCAGACCC CTCCCCAAAC ATGCATATGT ACCTGTCCGT
2801 CAGTGTGTGG ATGTATGTTT ACAGTTCTAC ATAAATGGGA TCATTTTATA
2851 CATGGTGCTC TGGAACCCAC ATTTTTCATG CAGTCATTG CAGTGAATTA
2901 TTTATGTGTA TAATAAATAG CATTAGAATA CAAAAAATAA AAAAAAATAA
2951 AAAAA (SEQ ID NO:2)

```

FEATURES:

5' UTR: 1-257

Start: 257

WO 02/059306

PCT/US01/42809

Stop: 2147
3' UTR: 2150-2955

HOMOLOGOUS PROTEINS:

Top BLAST Hits:

	Score	E
SPLICE FORM 1:		
gi 7302719 gb AAF57797.1 (AE003802) CG5002 gene product [Droso...	418	e-116
gi 7301881 gb AAF56989.1 (AE003772) CG7912 gene product [Droso...	408	e-113
gi 5834394 gb AAD53951.1 (AF180728) sulfate transporter [Droso...	388	e-107
gi 7301216 gb AAF56347.1 (AE003749) Esp gene product [Drosophi...	387	e-106
gi 7301962 gb AAF57068.1 (AE003774) CG9702 gene product [Droso...	351	8e-96
gi 7294633 gb AAF49971.1 (AE003543) CG6928 gene product [Droso...	347	2e-94
gi 7300023 gb AAF55195.1 (AE003708) CG6125 gene product [Droso...	330	3e-89
gi 7493011 pir T39116 probable sulfate permease - fission yeas...	267	2e-70
gi 6094367 sp O74377 SULH_SCHPO PROBABLE SULFATE PERMEASE SPBC3...	266	4e-70
gi 2626753 dbj BAA23424.1 (AB008782) sulfate transporter [Arab...	266	4e-70

SPLICE FORM 2:

gi 7302719 gb AAF57797.1 (AE003802) CG5002 gene product [Droso...	404	e-111
gi 7301881 gb AAF56989.1 (AE003772) CG7912 gene product [Droso...	394	e-108
gi 5834394 gb AAD53951.1 (AF180728) sulfate transporter [Droso...	374	e-102
gi 7301216 gb AAF56347.1 (AE003749) Esp gene product [Drosophi...	373	e-102
gi 7301962 gb AAF57068.1 (AE003774) CG9702 gene product [Droso...	337	2e-91
gi 7294633 gb AAF49971.1 (AE003543) CG6928 gene product [Droso...	332	6e-90
gi 7300023 gb AAF55195.1 (AE003708) CG6125 gene product [Droso...	315	8e-85
gi 7493011 pir T39116 probable sulfate permease - fission yeas...	254	2e-66
gi 9955547 emb CAC05432.1 (AL391710) sulfate transporter [Arab...	254	2e-66
gi 2626753 dbj BAA23424.1 (AB008782) sulfate transporter [Arab...	253	4e-66
gi 6502994 gb AAF14540.1 AF163975_1 (AF163975) Suta [Penicilliu...	250	3e-65
gi 6094367 sp O74377 SULH_SCHPO PROBABLE SULFATE PERMEASE SPBC3...	248	1e-64
gi 6502992 gb AAF14539.1 AF163974_1 (AF163974) sulfate permease...	244	2e-63
gi 10645530 gb AAG21641.1 AC069474_20 (AC069474) sulphate trans...	242	9e-63

BLAST to dbEST:**SPLICE FORM 1:**

gb AA582196.1 AA582196	nn44h12.s1 NCI_CGAP_GC5 Homo sapiens...	1027	0.0
gb AI690196.1 AI690196	tx33c05.x1 NCI_CGAP_Lu24 Homo sapien...	948	0.0
gb AI417381.1 AI417381	tg30c11.x1 NCI_CGAP_Brn25 Homo sapie...	902	0.0
gb AW207633.1 AW207633	UI-H-BI1-afl-c-10-0-UI.s1 NCI_CGAP_S...	890	0.0
gb AI494563.1 AI494563	qz16h03.x1 NCI_CGAP_CLL1 Homo sapien...	850	0.0
gb H15131.1 H15131	ym30b04.s1 Soares infant brain 1NIB Homo...	805	0.0
gb AI220943.1 AI220943	qg08c06.x1 Soares placenta 8to9weeks...	743	0.0
gb AW055140.1 AW055140	wz01c04.x1 NCI_CGAP_Brn23 Homo sapie...	718	0.0
gb AI963799.1 AI963799	wr67b09.x1 NCI_CGAP_Ut1 Homo sapiens...	718	0.0
gb N54913.1 N54913	yv34f02.s1 Soares fetal liver spleen 1NF...	710	0.0
gb AW008673.1 AW008673	ws71c02.x1 NCI_CGAP_Brn23 Homo sapie...	688	0.0
gb AI680222.1 AI680222	tw66e04.x1 NCI_CGAP_Ut3 Homo sapiens...	638	e-180
gb AA385773.1 AA385773	EST99540 Thyroid Homo sapiens cDNA 5...	636	e-180
gb H04038.1 H04038	yj45b06.r1 Soares placenta Nb2HP Homo sa...	626	e-177
gb AI000592.1 AI000592	os63b01.s1 NCI_CGAP_Br2 Homo sapiens...	618	e-174

BLAST to CHGI:

TA_39934	NOT ASSIGNED [Homo sapiens]	1544	0.0
TA_44508	NOT ASSIGNED [Homo sapiens]	509	e-142
TA_102843	NOT ASSIGNED [Homo sapi...	484	e-135
TA_82272	AAD14 [Homo sapiens]	44	0.018
TA_71453	ataxin SCAl [Homo sapiens]	44	0.018
TA_74619	alkaline phosphatase, pl...	42	0.070
TA_52146	NOT ASSIGNED [Homo sapiens]	42	0.070
TA_237492	NOT ASSIGNED [Homo sapi...	40	0.28
TA_158839	NOT ASSIGNED [Homo sapi...	40	0.28

EXPRESSION INFORMATION FOR MODULATORY USE:

SPLICE FORM 1:Library source from BLAST dbEST hits:

AA582196 534 bp mRNA EST 05-SEP-1997	germ cell tumor
AI690196 486 bp mRNA EST 16-DEC-1999	carcinoid:lung
AI417381 470 bp mRNA EST 30-MAR-1999	anaplastic oligodendroglioma: brain
AW207633 511 bp mRNA EST 02-DEC-1999	unknown
AI494563 457 bp mRNA EST 17-MAR-1999	B-cell, chronic lymphocytic leukemia
H15131 508 bp mRNA EST 27-JUN-1995	infant brain, 73 days post natal
AI220943 375 bp mRNA EST 29-NOV-1998	two placentae: one from 8 weeks and another
from 9 weeks post conception	
AW055140 366 bp mRNA EST 09-MAR-2000	glioblastoma (pooled): brain
AI963799 377 bp mRNA EST 08-MAR-2000	well-differentiated endometrial
adenocarcinoma, 7 pooled tumors: uterus	
N54913 486 bp mRNA EST 28-JAN-1997	fetal liver and spleen
AW008673 369 bp mRNA EST 10-SEP-1999	glioblastoma (pooled): brain

Library source from cDNA retrieval:

Human Fetal Brain
Human Brain
Human Pituitary Gland
Human Heart
Human Leukocytes
Human Kidney
Human Hela Cell
Human Liver
Human Thyroid
Human Lung
Human Placenta
Human Skeletal Muscle
Human Fetal Kidney
Human Small Intestine
Human Prostate
Human Testis
Human Adrenal Gland
Human Bone Marrow
Human Pancreas

SPlice FORM 1:

```

1  MPSSVTALGQ ARSSGPGMAP SACCSPAAL QRRLPILAWL PSYSLQWLKM
51 DFVAGLSVGL TAIPQALAYA EVAGLPPQYG LYSAFMGCFV YFFLGTSRDV
101 TLGPTAIMSL LVSFYTFHEP AYAVLLAFLS GCIQLAMGVL RLGFLDFIS
151 YPVIKGETSA AAVTIGFGQI KNLLGLQNIP RPFFLQVYHT FLRIAETRVG
201 DAVLGLVCM LLLVLKLMRD HVPPVHPEMP PGVRLSRGLV WAATTARNAL
251 VVSFAALVAY SFEVTGYQPF ILTGETAEGE PPVRIPPFSV TTANGTISFT
301 EMVQDMGAGL AVVPLMGLLE SIIVAKAFAS QNNYRIDANQ ELLAIGLTNM
351 LGSVLSSYPV TGSFGRTAVN AQSGVCTPAG GLVTGVVLVLL SLDYLTSLFY
401 YIPKSALAAV IIMAVAPLFD TKIFRTLWRV KRLDLLPLCV TFLLCFWEVQ
451 YGILAGALVS LLMLLHSAAR PETKVSEGPV LVLQPASGLS FPAMEALREE
501 ILSRALEVSP PRCLVLECTH VCSIDYTVVL GLGELLQDFQ KQGVALLAFVG
551 LQVPVLRVLL SADLKGQYF STLEEAEKHL RQEPGTQPNY IREDSILDQK
601 VALLKA (SEQ ID NO:3)

```

SPlice FORM 2:

```

1  MPSSVTALGQ ARSSGPGMAP SACCSPAAL QRRLPILAWL PSYSLQWLKM
51 DFVAGLSVGL TAIPQALAYA EVAGLPPQYG LYSAFMGCFV YFFLGTSRDV
101 TLGPTAIMSL LVSFYTFHEP AYAVLLAFLS GCIQLAMGVL RLAHISPHPL
151 GLGGAGTSSM SPLGWPGFLL DFISYPVIK FTSAAVTIG FGQIKNLLGL
201 QNIPRPFFLQ VYHTFLRIAE TRVGDAVLGL VCMLLLLVLL LMRDHVPPVH
251 PEMPPGVRLS RGLVWAATTA RNALVVSFAA LVAYSFEVTG YQPFILTGET
301 AEGLPVVRIP PFSVTTANGT ISFTMVQDM GAGLAVVPLM GLLESIAVAK
351 AFASQNNYRI DANQELLAIG LTNMLGSLVS SYPVTGSFGR TAVNAQSGVC
401 TPAGGLVTGV LVLLSLDYLT SLFYIIPKSA LAAVIIMAVA PLFDTKIFRT
451 LWRVKRLDLL PLWVTFLLCF WEVQYGILAG ALVSLMLLLH SAARPETKVS
501 EGPVLVLQPA SGLSFAMEA LREEILSRAL EVSPPRCLVL ECTHVCSIDY
551 TVVLGLGELL QDFQKQGVALL AFVGLQVPVL RVLLSADLKG FQYFSTLEEA
601 EKHLRQEPGT QPYNIREDSI LDQKVALLKA (SEQ ID NO:4)

```

FEATURES:

Functional domains and key regions:

SPlice FORM 1:

```

[1] PDOC00001 PS00001 ASN_GLYCOSYLATION
N-glycosylation site
294-297 NGTI

```

```

[2] PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
Number of matches: 2

```

```

1 96-98 TSR
2 245-247 TAR

```

```

[3] PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site
Number of matches: 6

```

```

1 96-99 TSRD
2 116-119 TFHE
3 298-301 SFTE
4 571-574 STLE
5 572-575 TLEE
6 595-598 SILD

```

[4] PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 15

1	9-14	GQARSS
2	17-22	GMAPSA
3	55-60	GLSVGL
4	80-85	GLYSAF
5	168-173	GQIKNL
6	232-237	GVRLSR
7	238-243	GLVWAA
8	307-312	GAGLAV
9	317-322	GLLESI
10	352-357	GSLVSS
11	380-385	GGLVTG
12	381-386	GLVTGV
13	452-457	GILAGA
14	456-461	GALVSL
15	543-548	GVALAF

[5] PDOC00370 PS01130 SULFATE TRANSP

Sulfate transporters signature

77-98 PQYGLYSAFMGCFVYFFLGTSR

Membrane spanning structure and domains:SPLICE FORM 1:

Helix	Begin	End	Score	Certainty
1	11	31	0.878	Putative
2	50	70	1.193	Certain
3	77	97	1.866	Certain
4	120	140	1.767	Certain
5	148	168	1.097	Certain
6	199	219	1.547	Certain
7	247	267	1.151	Certain
8	283	303	0.683	Putative
9	306	326	1.282	Certain
10	342	362	1.277	Certain
11	372	392	1.695	Certain
12	401	421	1.366	Certain
13	449	469	1.784	Certain
14	477	497	0.915	Putative
15	542	562	0.834	Putative

BLAST Alignment to Top Hit:

SPLICE FORM 1:

gi|7302719|gb|AAF57797.1| (AE003802) CG5002 gene product [Droso... 418 e-116

Query: 25 CSPAALQRRPILAWLPSYSLOWLKMDFVAGLSVGLTAIPQALAYAEVAGLPPQYGLYSA 84
C P+ + + PIL WLP Y L+++ DF+AG +VGLT IPQA+AY VAGL PQYGLYSA
Sbjct: 28 CRPSTVTNKFILKWLPRYRLEYIMQDFIAGFTVGLTTIPQAIAYGVVAGLEPQYGLYSA 87

Query: 85 FMGCFVYFFLGTSRDVTLGPTAIMSLVSYFTFHEPAYAVLLAFLSGCIQLAMGVRLRGF 144
FMGCF Y G+ +DVT+ TAIM+L+V+ Y P YAVL+ FL+GCI L +G+L +G
Sbjct: 88 FMGCFYIYVFGSCKDVTIATTAIMALMVNQYATISPDYAVLVCFLAGCIVLLGLLNMGV 147

Query: 145 LLDIFISYPVIKGFTSAAAVTIGFGQIKNLLGLQNIIPRPFFLQVYHTFLRIAETRVGDAVL 204
L+ FIS PVI GFT AAA TIG QI N++GL + + F + R+ DA+L
Sbjct: 148 LVRFISIPVITGFTMAAATTIGSAQINNIVGLTSPSNDLLPAWKNFFTHLTSIRLWDALL 207

Query: 205 GLVCMLLLLVLKLMRDHVPVHPPEMPGVRLSRGLVWA-ATTARNALVVSFAALVAYSFE 263
G+ ++ LL++ ++D ++ + W +RNAL V F +AY
Sbjct: 208 GVSSLVFLLLMTRVKD-----IKWGNRIFWKYLGLSRNALAVIFGTFLLAYILS 255

Query: 264 VTGYQPFILTGETAEGLPVRIPPFSVTTANGTISFTEMVQDMGAGLAVVPLMGLLESIA 323
G QPF +TG G+PP R+PPFS T +SF EM+ +GA L +PL+ +LE +A
Sbjct: 256 RDGNQPFVRTGNITAGVPPFRLPFFSTTVDGEYVSFGEMISTVGASIGSIPLISILEIVA 315

Query: 324 VAKAFASQNNYRIDANQELLAIGLTNMLGSLVSSYPVTGSEFGRRTAVNAQSGVCTPAGGLV 383
++KAF+ +DA+QE++A+G+ N++GS V S PVTGSF RTAVN SGV TP GG V
Sbjct: 316 ISKAFS--KGKIVDASQEMVALGMCNIMGSFVLSMPVTGSEFTRTAVNNASGVKTPLGGAV 373

Query: 384 TGVVLVLLSLDYLTSLFYIIPKSALA AAVIIMAVAPLFDTKIFRTLWRVKRLDPLPLCVTF 443
TG LVL++L +LT FY+IPK LAA+II A+ L + + +W+ K+ DL P VT L
Sbjct: 374 TGAFLVLMALAFLTQTFYFIPKCTLA AIIIAAMISLVELHKIKDMWKS KKKDLFFPVVTVL 433

Query: 444 LC-FWEVQYGILAGALVSLMLLHSAARPET-----KVSEGPVLVLQPASGLSFPAMEAL 497
C FW ++YGIL G +++ +L+S+ARP K++ V V+ L + + E L
Sbjct: 434 TCMFWSLEYGILCGIGANMVYIYSSARPHVDIKLEKINGHEVSVVDVKQKLDYASAEYL 493

Query: 498 REEILSRAL--EVSPPRCLVLECHVCSIDYTVVLGLGELLQDFQKQGVALLAFVGLQVPV 555
+E+++ R. L + + +V++ + SIDYTV + + + D + A+ +
Sbjct: 494 KEKVV-RFLNNQNGETQLVVIKGEINSIDYTVAMNIVSMKGDLEALNCAMICWNWNIA 552

Query: 556 LRVL---LSADLKG-FQYFSTLEE 575
V+ L+ DL+ F++ +LEE
Sbjct: 553 AGVVCRLNNDLRPIFKFDLSLEE 575 (SEQ ID NO :6)

SPLICE FORM 2:

>gi|7302719|gb|AAF57797.1| (AE003802) CG5002 gene product

[Drosophila melanogaster]

Length = 595

Score = 404 bits (1026), Expect = e-111

Identities = 233/587 (39%), Positives = 340/587 (57%), Gaps = 51/587 (8%)

Query: 25 CSPAALQRRPILAWLPSYSLOWLKMDFVAGLSVGLTAIPQALAYAEVAGLPPQYGLYSA 84
C P+ + + PIL WLP Y L+++ DF+AG +VGLT IPQA+AY VAGL PQYGLYSA
Sbjct: 28 CRPSTVTNKFILKWLPRYRLEYIMQDFIAGFTVGLTTIPQAIAYGVVAGLEPQYGLYSA 87

Query: 85 FMGCFVYFFLGTSRDVTLGPTAIMSLVSYFTFHEPAYAVLLAFLSGCIQLAMGVRLRLAH 144
FMGCF Y G+ +DVT+ TAIM+L+V+ Y P YAVL+ FL+GCI L +G+L +
Sbjct: 88 FMGCFYIYVFGSCKDVTIATTAIMALMVNQYATISPDYAVLVCFLAGCIVLLGLLNM-- 145

Query: 145 ISPHPLGLGGAGTSSMSPLGWPGFLLDFISYPVIKGFTSAAAVTIGFGQIKNLLGLQNIIP 204
G L+ FIS PVI GFT AAA TIG QI N++GL +
Sbjct: 146 -----GVLVRFISIPVITGFTMAAATTIGSAQINNIVGLTSPS 183

Query: 205 RPFFLQVYHTFLRIAETRVGDAVLGLVCMLLLLVLKLMRDHVPVHPPEMPGVRLSRGLV 264

+ F + R+ DA+LG+ ++ LL++ ++D ++ +
 Sbjct: 184 NDLLPAWKNFFTHLTSIRLWDALLGVSSLVFLLLMTRVKD-----IKWGNRIF 231
 Query: 265 WA-ATTARNALVVSFAALVAYSFEVTGYQPFILTGETAEGLPVRIIPFVSVTANGTISF 323
 W +RNAL V F +AY G QPF +TG G+PP R+PPFS T +SF
 Sbjct: 232 WKYLGLSRNALAVIFGTFLAYILSRDGNQPFRTVGNITAGVPPFRLPPFSTTVDGEYVSF 291
 Query: 324 TEMVQDMGAGLAVVPLMGLLESIAVAKAFASQNNYRIDANQELLAIGLTNMLGSLVSSYP 383
 EM+ +GA L +PL+ +LE +A++KAF+ +DA+QE++A+G+ N++GS V S P
 Sbjct: 292 GEMISTVGASLGSIPILISILEIVAISKAFS--KGKIVDASQEMVALGMCNIMGSFVLSMP 349
 Query: 384 VTGSFGRRTAVNAQSGVCTPAGGLVTGVVLVLLSLDYLTSLFYYPKSAALAAVIIMAVAPLF 443
 VTGSF RTAVN SGV TP GG VTG LVL++L +LT FY+IPK LAA+II A+ L
 Sbjct: 350 VTGSFTRTAVNNASGVKTPGGAVTGALVLMALFLTQTFYFIPKCTLAIIIAAMISLV 409
 Query: 444 DTKIFRTLWRVKRLDLLPLWVTFLLC-FWEVQYIGILAGALVSLMLLHSAARPET----- 497
 + + +W+ K+ DL P VT L C FW ++YGIL G +++ +L+S+ARP
 Sbjct: 410 ELHKIKDMWKSKKKDLFPFVTVLTCMFWSLEYGILCGIGANMVYILYSSARPHVDIKLE 469
 Query: 498 KVSEGPVLVLQPASGLSFFAMEALREEILSRAL--EVSPPRCLVLECTHVCSIDYTVVLG 555
 K++ V V+ L + + E L+E+++ R L + + +V++ + SIDYTV +
 Sbjct: 470 KINGHEVSVDVKQKLDYASAEYLKEKVV-RFLNNQNGETQLVVIKGEINSIDYTVAMN 528
 Query: 556 LGELLQDFQKQGVAFVGLQVPVLRVL--LSADLKG-FQYFSTLEE 599
 + + D + A+ + V+ L+ DL+ F++ +LEE
 Sbjct: 529 IVSMKGDLEALNCAMICWNWNINIASAGVVCRLNNDLRPIFKFDLSLEE 575 (SEQ ID NO :7)

Multiple sequence alignment of both splice forms and the top blast hit:

Splice Form 1 ~~~~~~ --MPSSVTAL GQARSSGPGM APSACCCSPA ALQRRLPILA
 Splice Form 2 ~~~~~~ --MPSSVTAL GQARSSGPGM APSACCCSPA ALQRRLPILA
 gi7302719_pe MRAEDNLYR EQLPNVSTLI ...RDGGRKLCRPS TVTNKFPILK

51 100
 Splice Form 1 WLPSYSLQWL KMDFVAGLSV GLTAIPQALA YAEVAGLPPQ YGLYSAFMGC
 Splice Form 2 WLPSYSLQWL KMDFVAGLSV GLTAIPQALA YAEVAGLPPQ YGLYSAFMGC
 gi7302719_pe WLPRYRLEYI MQDFIAGFTV GLTTIPQAIA YGVVAGLEPQ YGLYSAFMGC

101 150
 Splice Form 1 FVYFFLGTSR DVTLGPTAIM SLLVSFYTFH EPAYAVLLAF LSGCIQLAMG
 Splice Form 2 FVYFFLGTSR DVTLGPTAIM SLLVSFYTFH EPAYAVLLAF LSGCIQLAMG
 gi7302719_pe FTYIVFGSCK DVTIATTAIM ALMVNQYATI SPDYAVLVCF LAGCIVLLLG

151 200
 Splice Form 1 VLRL.....GF LLDFISYPVI KGFTSAAAVT
 Splice Form 2 VLRLAHISPH PLGLGGAGTS SMSPLGWPGF LLDFISYPVI KGFTSAAAVT
 gi7302719_pe LLNM.....GV LVRFISIPVI TGFTMAAATT

201 250
 Splice Form 1 IGFGQIKNLL GLQNIIPRPF LQVYHTFL.R IAETRVGDAV LGLVCMLLLL
 Splice Form 2 IGFGQIKNLL GLQNIIPRPF LQVYHTFL.R IAETRVGDAV LGLVCMLLLL
 gi7302719_pe IGSAQINNIV GLTS.PSNDL LPAWKNFFTH LTSIRLWDAL LGVSSLVFLL

251 300
 Splice Form 1 VLKLMRDHVP PVHPEMPPGV RLSRGLVWA. ATTARNALVV SFAALVAYSF
 Splice Form 2 VLKLMRDHVP PVHPEMPPGV RLSRGLVWA. ATTARNALVV SFAALVAYSF
 gi7302719_pe LMTRVKD...I KWGNRIFWKY LGLSRNALAV IFGTFLAYIL


```

301                                     350
Splice Form 1 EVTGYQPFIL TGETAEG LPP VRIPPF SVTT ANGTISFTEM VQDMGAGLAV
Splice Form 2 EVTGYQPFIL TGETAEG LPP VRIPPF SVTT ANGTISFTEM VQDMGAGLAV
gi7302719_pe SRDGNQPFVR TGNITAGVPP FRLPPFSTTV DGEYVSFGEM ISTVGASLGS

351                                     400
Splice Form 1 VPLMGLLESI AVAKAFASQN NYRIDANQEL LAIGLTNMLG SLVSSYPVTG
Splice Form 2 VPLMGLLESI AVAKAFASQN NYRIDANQEL LAIGLTNMLG SLVSSYPVTG
gi7302719_pe IPLISILEIV AISKAFS..K GKIVDASQEM VALGMCNIMG SFVLSMPVTG

401                                     450
Splice Form 1 SFGR TAVNAQ SGVCTPAGGL VTGVLVLLSL DYLTSLFYI PKSALAAVII
Splice Form 2 SFGR TAVNAQ SGVCTPAGGL VTGVLVLLSL DYLTSLFYI PKSALAAVII
gi7302719_pe SFTR TAVNNA SGVKTPLGGA VTGALVLMAL AFLTQTFYFI PKCTLAAIII

451                                     500
Splice Form 1 MAVAPLFDTK IFRTLWRVKR LDLLPLCVTF LLC.FWEVQY GILAGALVSL
Splice Form 2 MAVAPLFDTK IFRTLWRVKR LDLLPLWVTF LLC.FWEVQY GILAGALVSL
gi7302719_pe AAMISLVELH KIKDMWKSCK KDLFPFVTV LTCMFWSLEY GILCGIGANM

501                                     550
Splice Form 1 LMLLHSAARP ET....KVS EGPVLVLQPA SGLSFPAMEA LREEILSRAL
Splice Form 2 LMLLHSAARP ET....KVS EGPVLVLQPA SGLSFPAMEA LREEILSRAL
gi7302719_pe VYIYSSARP HVDIKLEKIN GHEVSVVDVK QKLDYASAAY LKEKVV.RFL

551                                     600
Splice Form 1 ..EVSPRCL VLECTHVCSI DYT VVLGLGE LLQDFQKQGV ALAFVGLQVP
Splice Form 2 ..EVSPRCL VLECTHVCSI DYT VVLGLGE LLQDFQKQGV ALAFVGLQVP
gi7302719_pe NNQNGETQLV VIKGEEINSI DYT VAMNIVS MKGDLEALNC AMICWNWNIA

601                                     650
Splice Form 1 VLRVL..LSA DLKG.FQYFS TLEEA EKHLR QEPGTQPYNI REDSILDQKV
Splice Form 2 VLRVL..LSA DLKG.FQYFS TLEEA EKHLR QEPGTQPYNI REDSILDQKV
gi7302719_pe SAGVVCRLNN DLRPIFKFDL SLEEVVAGHF DSPSNTASTV TIEA~~~~~

651
Splice Form 1 ALLKA
Splice Form 2 ALLKA
gi7302719_pe ~~~~~

```

Hammer search results (Pfam):

Model	Seq-from	Seq-to	HMM-from	HMM-to	Score	E-value	Description
Sulfate_transp family	137	468	1	328	229.9	1.3e-66	Sulfate transporter

1 GGTCCCCGCG GCCCTCGGCC TTGCTCGGGG CCAAGGGACC GCGGACGGTC
51 AGGTGGCGCA GGGTCTCCTC CGGAGACCCC AGGATCCGGA GCCAGCGGCC
101 TTGTGGGCAG GGGCCGGGGG CAGGGGAGTG GATTTTGCCC GGAGCGGAGC
151 AGGCCGGGGG CAGTGGGGGG CTGGGGGTGA GGGTGGCTGG CTCTGCGCGC
201 GGGCGCCGGG GCCCTGGAAG ATGCTGCGCA CCTGAATTAA CCGGGCGCCT
251 CTGATGTCCT CCCAGAAGCA ACTAGAACTC CAGGGCTGTG AAAGCCACAG
301 GTGGGGGCTG AGCGAGGCTT GGCTCAGGA GCGGAGGACC CCCCCCCCCC
351 CCCCCTCGAG CGCCGCACTC CACCGTAGCG GGTGGAGCCC GCCTTGGTGC
401 GCAGTTGGAA AACCTCGGAA GCCCCGCTGG ATCTCCTGGC TGCCACCCGC
451 ACCCCCCGCC AGCTACGGTG CGCCCCGGG CCCAGCTTCT CTCTGCGCTG
501 CTCCCCGTTA AATTCCCTGG GGAGACGGAA AAAAAGGCAA AGGAAGTCGG
551 TTCTCCAGGG GCCAGAAGTG TTGAGCCTAA TTAGTCTTCA GACTTCTCAA
601 TGAGGAATCG CTTATCAGTT TCTTATCTGG GAGAGTTGAG GATGGAGGGA
651 CAGAAGGCAC CCAGGATTG CACGGGGGGG GATTACAGGA GAGAGGGTGA
701 TGAGGGACGG GGTGGGCCTT CCAGTCTTGG CCCAGTCCCC ATCTGCACA
751 CATTGTTGGC TTCTCTTAG AGCCGTTTCG CCCCCTGGGG AGGGGAGACC
801 CATAGTGACC TCTCTGACA CCCGCCGACC CTGACCAGTG TTGCCGGGT
851 CTTCAAAGGC CACGCTCTGA CTGCTGGTCT GTGTACCTG CACCCCCAG
901 CCCCACCGTA GAGATGCCTT CTTGGTGAC GCGCTGGGT CAGGCCAGGT
951 CCTCTGGCCC CGGGATGGCC CCGAGCGCCT GCTGCTGCTC CCCTGCGGCC
1001 CTGCAGAGGA GGCTGCCCAT CCTGGCGTGG CTGCCCAGCT ACTCCCTGCA
1051 GTGGCTGAAG ATGGATTTCG TCGCCGCCCT CTCAGTTGGC CTCACTGCCA
1101 TTCCCCAGGC GCTGGCCTAT GCTGAAGTGG CTGGACTCCC GCCCCAGGTG
1151 AGGCGTCTGA CCCTGCTGCC AGCCATATCT CAGAAACAGT GCAGAATACA
1201 CAGTATCAAT CCCAGACACC ATCAGCGATT CCAGGTTTCC AGCCCCTGGG
1251 CCCCAGGAA CCTTTGGTTT ACAGTGTGTG ACGCAGATTG TCTCTGGGCC
1301 GACCCAGGCT CCTATGCCTG TTTGGTACAC ACAGACACTG AGCTGGTTAT
1351 GGAGGGGCCA GCGAGATGAC TCATGGAGGC CTCAGGAGTT CAAGACCAGC
1401 CCGACCAAAA TGGTGAAACC CCGTCTCCAC TAAAAATACA AAAATTAGGC
1451 TGGGTGCGGT GGCTCAAGCC TGTAATCCCA GCACTTTGGG AGGCCGAGGC
1501 AGGCGGATCG CAAGTCCAGG AGATGGAGAC CATCCTGGCT AACAGGGTGA
1551 ACCCCGCTCT CTAATAAAAA TACAAAAAAT TAGCCAGGTG TGGTGGCGGG
1601 TGCTGTAGT CCCAGCTACC TGGGAGGCTG AGGCAGGAGA ATGGCGTGAA
1651 CCGGGGAGGC GGAGCTTGCA GTGAGCCGAG ACTGTGCCAC TGCCCTCCAG
1701 CTTGGGCGAC AGAGCGAGAC TCCATCTCAA AAAAAAAAAA AAAAGAAATGC
1751 TTCTCAGAC TTGGACACAG CACACGGGCC TGTAACGACC CCTCTGCCTG
1801 GCTGTCTGCA CCCTGAGGCC CCAGTTGAGT GCTGCTAAAA AAGTGGCCTC
1851 CTGATCACTG CAGTCCACC CACAGGGCAG GGCGGTGCAC CTTTAACCTG
1901 GGCTGGACA CAGTGACAC CCACACATCC CGAGCTTGA CACGCACACT
1951 AGGGAGCTGG TGGATGGGCC TCGGCCTCCT GAGTGTCTAC CACCCTCTCT
2001 CCCCACAGTA TGGCTCTAC TCTGCCTTCA TGGGCTGCTT CGTGTATTTT
2051 TTCTTGGGCA CCTCCCGGGA TGTGACTCTG GGCCCCACCG CCATTATGTC
2101 CCTCTGGTC TCCTTCTACA CCTCCATGA GCGCGCTAC GCTGTGCTGC
2151 TGGCCTTCCT GTCCGGCTGC ATCCAGCTGG CCATGGGGGT CCTGCGTTTG
2201 GGTGAGGCTC TACCTTCTTG CCAAGGGGAT GCCCTCGACC TCAGCATTTG
2251 CTTGTTTGCA TTCAAGTCT ATCCCCGTGT GCGTGTGTGT GCGTGTGTG
2301 GGGGTGTGGG TATGTATGTG TGTGTGTGTA GGTGGGTGGG TGGTGGAGGG
2351 GGTGGGGCAC TTGGCTCCTT AGTCTACTAT TTTACTGATT AGAGGCCAGG
2401 ACATTGGAGA AAGTGACCTG TGGCTCAGAC CCCATATGCC CCNNNNNNNN
2451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNTTCATTTT
2501 CTACCCCGTC ATTAAAGCTT CACCTCTGCT GCTGCCGTCA CCATCGGCTT
2551 TGGACAGATC AAGGTAGGCA CGGCGCCAC CCAGGGCACT GCTCTTTGGC
2601 CACTGCTCGT TGGCACAGGG ATGGCGGGAG CAGGACTGAG GCCAGTCCTG
2651 ATCCCTGTGG CCAGTGGACG TCTTGCTGTT TCAGATTGTC TTCCATGGGT
2701 CAAGAAGCAC GCGGTGCTCT CATGGGTCCC CTGTTAATAA AATGACCCCTC
2751 CTGGGAGGGA TGTCACGTGA TGGTTGGATT TCACAGCGGG TAACTTGGGG
2801 GCGGTAATT CCATCCCCCT GCTCTTGCCC GAGTTTCCGT GCCAGTGTGC
2851 TTGGCTGGCT CTGTGACGTG GCTCTGTTCT CCCTGCACTG GGCACACCCA
2901 GCAGGCCCCA CCAGTCATGA GCATGCTGCT AGAATTTCTA TAGGCAAATT
2951 ATTTCCCAT GCCAATTTAG TTAGATGGTT TTGTTTTGTG CTTTTTTTTT
3001 TGGTTTCTGT TTAATATTTT TTAATGCCA TGCTTTTAT ATTTTCTTTT
3051 AAAACATTTA TAATAAGTAC TAAATCAGT CATCAGCTGA GGGTGAATT
3101 TATTTCTGTT TTGCTGGGTT GTAAGTTCTT CGAGGGTAGG AACTGTCCTT
3151 CTCCAGCCA TGGCCGCCCA GCATTGGGCT GTGTCAGTAG GGGGGTGTCT
3201 AGTGGGGTGT GTGTGGAGTG AAGTGAGCCC AGTCCACAG ATGGGACCAT
3251 GCGGCCTCA TGGCAGACTA GGGTCACATG CTGCCTCTG ACCCTGTGTC

```

3301 ACTGCAGGTC ATCTTCCCCA AGCCAGGCCC TGTTCAGGC TGGCCTGAGA
3351 CAGTCTTCCC TGATGGAGGT ACCATGAGAA GACCAAGGAC AGGAGAGTGT
3401 GTGTGAGAGT GTGTATGAGT GTGTGTGTGT GTGAATGAGT GAGTGTGTGA
3451 GAGTGTGAGT GGGTTTGAGG GAGTGAGTGT GTGCGTGTGT GAGTGAGTGT
3501 GAGTGTGGGT GTGAGTGAGT GCGTGTGTGT GAGAGAGTGT GTGAGTGTGT
3551 ATGAGTGAGA GTGTGAGAGT GGGTGTGGGT GTTAGGGAGT GTGTGAGTGT
3601 GCGCGCGCGT GTGTGAGTGT ATGAGTGTGA GAGTGAGTGT GAGTGAGTGA
3651 GAGTGGGTGT GGGTGTGAGT GTGCGTGTGA GTGTGAGAGT GTGAGTGTGT
3701 GAGTGTGAGA ATAAAGTAGA CACTTTTTGC ACTCTTGCTA CGTGCAGAGC
3751 ACTGGGCGAGG ACACTCTCCA TGTGTAATTC TCAACACACC CCGGGAGGTA
3801 GATTTATCAT TATTTCCATT TGAGAGAGGA GGGACCAACT TAGGTGTGGG
3851 TGAGCGTGAT TTGTACGTTA TCTACACGCA TCTCTCAGGG TTAGCTTGGC
3901 AAATGCTGTT TCAGGGCATG GTTGGTTCTT TAATCTGGAA ACATCATTTT
3951 TGGTGTCAAG AATGTTCTTT TGTAGGATCC CAGTGAGAGT GGAGAGCGGA
4001 GAGTGGAGAG TGAAGGCAT CCCTTGTTCA TGCCTTCATA CTTGGCAACC
4051 CTAGCCCCGC CCAGGGACTC TGCAGCCATC TGGGGGGAGG GCGTCTCTCC
4101 TGACAGGCCC AGGACAGAAG ACCCTACCCC AACAGTCCCA GAGTCCGCCC
4151 CCCAGGATGT CCTAACCCCA CCCACCCTGT ACGCAGCAGT TTTAGGGCAG
4201 GGTCTTGTAG CCTGTGGCCA TGGGATCCAG GGCCTCAACT TTCCCTCTG
4251 ACTCCTGTTG GGTCTCAGGC GATGTAAAGA ACTAAAGTGC AAGCTGTGCC
4301 TGCAGTGTGG ATCCCAGCTA CGTGGGAGGC CGAGGTGGGA GGATCACTGG
4351 AGCCCAGGAG TTCGAATCCA GCCTGGGCCA CAGAGCAAGC AAGACCCCA
4401 CCCCCAACCC CCCGCGCGCT CTCTAAAGAA AAAAAACTA AACTAAACA
4451 CAGGATAGAG TGCTTCTCC TGCCAGGAC CTCAGAGCTG GTATCGTGGT
4501 GGGAGGCTCC TACTTTGCCG AGGATTCCCC AAGCTGGTTT CTTGAAGCCC
4551 CTCAGAGCCC TCCACATCTG CACACTACGA AGAGATTTT CCTCCGCGAG
4601 CAGCGGAGCT GGGGGGTGGC GGGGCACCTA GTGAGGGAGA CATTCTCAAG
4651 CCAATGGCAG CAAGGGTCTA CACTGCAGGG GGCCTGGCTG CTGAGTGGCC
4701 TGGAGCCAAT GGGGGTGGGG CCAGGCAGCC CATGGCCCTT GGCCATCAAG
4751 CTGTAGAACC TGTCTGCCTG CTGTGGGCTG ACCTACATTG TTTTATGA
4801 GAAAAGTAAT TTAGAGAAA ACATATCACT GACCCAGTAA TTGTGAAGTA
4851 CTGTCTCCA CGAGGGTAGC TTTGATCTCC TGCCCTAGGG GCGTTGGGA
4901 GTGGGCAGGT GGACCCCTGA GCCCTAAGA TGGGCCCCGG GAATAAGGGT
4951 TGGGAGCAGG GCCGGGGGAC ACTGTCTCAG CCCTAGGGGA GGTGGGCGGG
5001 GAGCTGGGGA CAGATGGCCT TGGTTGGGA GCATAGCCTC TGATCAGCAT
5051 CTCTGTGTTT GGACAGAACC TGCTGGGACT ACAGAACATC CCCAGGCCGT
5101 TCTTCTGCA GGTGTACCAC ACCTTCCTCA GGATTGCAGA GACCAGGTAC
5151 CCCGGGCTTT GTTCTCCTT CCTATAAGGA AGCTCCTTCT TCCACACCTC
5201 CTCTCCCGGC CCCACCTCA GTTCCCCAC CCCTGGTGAC TGCTCAAACA
5251 GGGGTCCCCA GAGCAGCCCC ATCAGCAGCA CCTGCAAGCT GGCAAGAAAT
5301 GAGACTCTCA GTCCTTTGCA GACCTGCCCT ATCGGAGCCA GAGGTTAAC
5351 AGGAGGCCGG TGGTCTGCGC GCAGGGGAGA GTTCAAGAAG CTCTGCTCTG
5401 GGGAGCTGGA GCAGGGCAAT GTTCTCTCT CCACAAAGCT TTCTTGGGG
5451 ACAGGAATGG GGAGAGTCTG TGTAAAGAAA AAGAAGGCAG TGTGTCTCCC
5501 CCAGGCTGTG ATTTGTTAAG GAGGAGAACA CAGGGCGTGG GGAGCTAACC
5551 CAGACAGAAC GCAGGTGGTG CAGGGACGGC AGGTGGAGCT GTGATGAGAG
5601 ATGAGGAGAC CCAAGCACTT CCCTCAGGTG GTTTAAAGAG CCTCTGATGG
5651 CCGGACACTG TGGCTCACAC CTGTAATCCC AGCACTTTGG GAAGCCAGG
5701 TGGGCAGATC ATCTTAGGTC AGGAGTTCAA GACCAGCCTG GTCAACATGG
5751 TGAAACTCCG TCTCTACTAA AAATACAAAA ATTAGTCGGG TGTGATGGTG
5801 CGTGCCTGTA GTCCAGCTG TTCAGGAGGC TGAGGCAGGA GGATTGCTTG
5851 AACCTGTGAG GCGGAGGTTG CAGTGAGCTG AGATCACGCC ACTGCACCTC
5901 AGCCCGGGTG ATAGAGTGAG ACTCCATCTC GCCTTGCCTG TGCCCCCGG
5951 AAAAGAAAAG CCTCTGATGA GGGGTACCTC CCTGCCAGAC CATCCAGCGG
6001 GAAGGCAGGA TGCCTCTCTA CCTCTCTGGC TGGAAAGGGC TGGAGGAGGA
6051 AATGATCTAG GGGAGCTATA GAGATGGCTG CCCAGTGCTG TGGCCTGGAG
6101 GGAGTGGGGG CGATCCTGGA CCATCTTCCC CTCCCTCTG ATCTGCTGCC
6151 GAAGCTTCCC AGCAGCTGAA CCCAGCTGGA AGCCAGTGGG TTCTTTGCTC
6201 TTCAGAGGCA CCAGTGGGCA GGGGTGCGCC AGATGGGGGA GCAGTGATTA
6251 CGGAGCCTGA CCAGTCAGAG AGCGTCACGG GCTGTCTATG AAAAGGGGCC
6301 TTAGTGACCG TTTAGTCCTA AGCCTTCTAG AATGTCTAGA GCAGGGTGTC
6351 CAGTCTTTTG TCTTCCCTGG GCCACACAGG AAGAATTGTC TTAAATACA
6401 CTAATGATAG CTGATGAGCT ACAAAAAAAT CATTGAAAGA ACTCATAATG
6451 TTTTAAGAAA GTTTACAAAT TTGTGTTGGG CCACATTCAA AGCCATCCCA
6501 GGCTGCACAT GGCCACAGG CTGTGGGTTT GGCAGACTTG TTCTAGAGCC
6551 ACCTGCAGAG AGCATCAGCT GCTCAGGCTG GTCCCCATTG CCTGTGGTCA

```

6601 CTCACCAGAT CCTGTTCCAT AGACTTGAGT CCCAGAGAGC TACGGGAGTG
6651 AAAATGTGAG CACGTCGCTC TGCCATCATC ATGACCCGGG CCTCTCCACT
6701 GCCTGTCCTG CCCAAGCCTG TGTACTTACT GAATTTTGAA TTGAGTTTGT
6751 AAGTAGAACA GGGATGCATA TTGCTTAGGA GAAAACATCT CAGCCTAGTC
6801 TGTGTTTGGT CTTGCAGAGT AGCTTGCTAA AGTTCCCTGAG CTTTAGTTAA
6851 GTGTCTGTGA AATGGTGAAA ATATTGAAAA TGCTTTATTT CCTTGTGATA
6901 CTCACTGTGG TCTGGGACCA GCAGCCTGGG CATTGGGCCT CCCTGGGAGC
6951 TTCTTAGTGC CTCTCAGCTT AAGCAGCAGC AGGACGCTTG AGTCAGCAGT
7001 AGGCCTGGGC CGAGCTGGGT GGTGACCAGT CCTCTGCCTG TCCACAGGGT
7051 AGGTGACGCC GTCCTGGGGC TGGTCTGCAT GCTGCTGCTG CTGGTGCTGA
7101 AGCTGATGCG GGACCACGTG CCTCCCGTCC ACCCCGAGAT GCGCCCTGGT
7151 GTGCGGCTCA CCGTGGGCT GGTCTGGGCT GCCACGACAG GTGAGGGGCC
7201 TCTGGCTGAC ATCGTATGCA ACCTGGGCTG CAGGTGGGG TCACTGGGG
7251 AGTCCTAGTC CCACCCTAGG GATTCTCAGC TCATTGGTCT GGGTGTCACT
7301 TGAGCATTGG GACATTTAAA ACACCACACC AAACCTCTGG ACATGTACTT
7351 TTTATTTAAT TAATTAATTA ATTATTTTTT GAGACAGAGT TTCGCTCTGT
7401 CGCCCAGGCT GGAGTACAGT GGTGTGATAT TGGCTCACTG CAACCTCCGC
7451 CTCCCAGGTT CATGTGATTC TCCTGCCTCA GCCTCCTGAG TAGCTGGGAT
7501 TACAGGCGCA TGCCACCACG CTCGACTGAT TTTTGTATTT ATAGTAGAGA
7551 CGGGGTTTCA CCATGTTGAC CATACTGGTC TCAAACCTCT GACCTCGAGT
7601 GATCTGCCGT CCTCGGCCTC CCGAAGTGCT GGGATTACAG GCATGAGCCA
7651 CCGCCCGGAC CCAAACCTCTA GAACATTTAA AACTCTGAGC CAAACTCGGT
7701 GGGTTCGTAT GCAGGGTCAG AGCTGAGAAC TGATGCATGG CCAGCTGTGG
7751 CCGTTCCTCA CCTGAAGCAG TTTTGCCCCC CTGTTGACAT GTGGCAATGT
7801 TGGGAAACAT TTTTGGTGGA TACCACCGGG CGAGGCAGTT GCTGGCAACT
7851 AGCGAGGAGG GGCCAGGAAT GCCACTAAGC CCCTGACGGC GCACAGCCTT
7901 CCACAGCAAA CAGTGATCCG GCCCAAAATG TCGGCGGTGC CGAGGTGGAG
7951 AAACACAGTC TACTGGCAGG TCCTGGGGAG ACAGAACAAA TTCCGGGGAG
8001 ACCAACAGGC TCTGTTTGAA TTGTGCAGCT TTATGGGGGT ACAACTTCAG
8051 CTTAGGCGC GGAGGCTGCA GGCTGAGCCA CAACCAACCA GGGGTCTTT
8101 GAGAAAGGGG TGGTGAGGGC TCGTCTGTGG AGGAGCAGGG CTGTCAGCAG
8151 GTGAAGGACC ACCCTGGGGG AGGTGGAAGC CTCTTCTCAG TGCATCCGGG
8201 ATGGGCATGG GTCCTCTGGC TGAGCACCGG GCAGGGATGC CCAGAGAAGA
8251 AGGCACACCC AGGAAAGTCA GGAAGCATAG GTGACTCCAA ACCTTCTCAG
8301 AGCCGAGGGG ACAGATTAAG CTCAGAATAA CCGAGCTTG GTCTTGCTGT
8351 TTGGGGGCTG CAAAATCCAA CAGCCACAGA AGAGAGAGGG TGGCTGGATT
8401 TAGCTCTTGA GTCACCTCTG CCTGCCCAGG CTGGTCTGTT CTCTGTGGTC
8451 TCTCCCTGGA TCCTGTTTCA CAGACTTGAG GCCGTGGTGA AACTGAGAG
8501 TCTGGTGGGG TGCAGAGGAT GCTGGGTTA GAGTGGAGAT GTCCAGCTTC
8551 CAGTCCCAGC CTGGCCTGGG TTAGACTTG CAAAGTAGCT TCCAAAGTTC
8601 CTGAGCTTTA GTTCAGTTGT CCGTGAAATA GTGAAATGC TTTGTTATCT
8651 TTGTATCTT TTGTTTTTTG TTTTTTTTTG GAGACGGTGT CTCACCTCTG
8701 TTGCCCAGGC TGGAGTGCAC TGGCCGTGAT CTTGGCTCA CTGCAACCTG
8751 CCTCTGGTT TTCAAGCAAT TATCCCCC TTAACCTCC CGGAGTAGCC
8801 TGGGATTATC AGGCCGTGCA CCCACCACAC CCGGGCTAAT TTTTGTATT
8851 TTTGAGTTCA AGATGGGGT TTCACCATTG TTGGCCCATG CTGGGTCTCG
8901 AAACCTCCCN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
8951 NNNNNNNNCC CTAACCTGTA AAGGCAGTGT CTCCGACGCC AGAAAGGTCT
9001 TTTAAGATGT TTAGACATCG GTAATGTCCA GAAACTTCAG AAACATTTTA
9051 CAGTGTAGGG CTGCTGACTC CTTTCTCCCC AGTGTGTGTA GCATTGTCTT
9101 AGGAAACTAC TTAATAAGCT TTTGATTCTT ATAACAGCTT TGAGAATGAG
9151 TTTATTATTT TTTCCATTTT GTCAGTGAGG AAAGTGAGCC TATGAAAGAT
9201 GAAAAATAAA ACAGAACTTC TCAAGGCCAC ACGACTGGGA AGAGGCAGAA
9251 CAAGGACTTC ACCCGAGTCG TCTGTCTTTT TTTCTGAATG GTCTGTCTAG
9301 AAGTGTGTCA ATGTTATGAT CCTCTCAAAG AACCAGCTTT TGGTTTCATT
9351 GATCTCCCTT TTGCTTTGGT TTTCTAGTTC ATTGATTCTT GCTTGGGCCT
9401 TTATTATTTT CTTTCTTCTA TTTATTTTCA GTTAACTTTA TCCTTTTTTT
9451 TTTTTTTTTT AAAAAAAAT TTTTTTTTTT GAGACGGAGT TTCGCTCTTG
9501 TCCCCCAGT TGGAGTGCAA TGGTGTGATC TCAGCTCACT GCAACCTCCG
9551 CCTCCCGGGT TCAAGTGATT CTCCTGCCTC AGCCTCCCAA GTAGCTGGGA
9601 TTACAGGCGT ATGCCACCAT GCCTGGCTAA TTTTTGTAT TTTAATAGA
9651 GACGGGGTTT CACCATGTTG GCCAGGCTGG TCTCGAACTC CTGAACTCAG
9701 GTGATCCACC CGCCTCAGCC TCCCAAAGTG CTGGGATTAC AGGCGTGAGC
9751 CACTGTGCCC GGCCTTTTTT TTCAGATTCT TAAGTGGGAA GCCTGGGTCT
9801 TTGATTTGAG ACCTTCTTCT TTTTCTAATA TGGTGTGTTA AATGCTGTAA
9851 ATTTCACTCC TAGCATTACT TTAGCTGCAT CCCACAGTTT TTATATATAT

9901 GCTGTTCTTT AATTTTATT CGGTTCTAAA TACTTTGTGA TTCCATTTT
9951 GACTTCTCCT TTGACCTGTG GATTATTTAG AAGTGGATTA TATGATTCT
10001 TCTGTGAAGC ATGGGAGTTA TTTAGTTTCC AAATAGTTGG GGGTTTCCCA
10051 ATTAGTACTT TGCCATTGAT TTCTAATTTA GTTCCATTGT GATAAGAAAA
10101 CGTACTTAGT GATTTAAATC TGTTTTCTCT TATTGAGAGT TGTATTATGG
10151 CCCTGAATAT GGTCTATCTT AGTGAATATT CTGTGTGTGC TTGAGAACAT
10201 GTTTTCTGCC ATTATTGGGT GGAATGTTCT GTAAATGTCA GTTAGGTCTG
10251 GTTGGTTATC TCAGGCTACT ATATTCTTGC TGATTTTCTT TCTACTTGT
10301 CTGTTGATTA TTAAAAGAAG AATGTTGAAA TAACTTCAGC TATAACCATG
10351 ACTTTGTCTA TTTTCTTTC TTACTTACCA GTTATTGTTT CATATATTTT
10401 GAAGCTCTGT TTTTGTATAT AAAACATAT AGAATTATAA TTTTCTCTTG
10451 GAATGATACT TTTTCTCTG GGCAATATTC TCTTCCCTGA AATCTACTTT
10501 GATACATAAGA GAGCTACTCC AGCTTTCATT AGTGTTAGCA TATCTTTTCC
10551 CATCTTTTTA CTTTTTTTTT TTTGAGACAG AATCTCTCTC TGTGCCCAG
10601 GCTGGAGTGC AATGGCGTGA TCTTGGCTCA CTGCAACCTC CACCTCCCAG
10651 GTTCAAGCAA TTGTTGTGCC TCAGCCTCCC AAGTAGCTGG AATTACAGGC
10701 TTGCACCACC AAGCCAGCT AATTTTGTG TTTTGTAGT CAATGAGGT
10751 TCACCATGTT GGCCAGCCTG GTCTCAAATT CTTGGCCTCA AGTGATCCAC
10801 CTGCCTTGCC CTCCCAAAGT GCTGGGACAC CACACCCAGC CTCTATCTT
10851 TTTACTTTTA ACGTATTTAA TTGATTGTG TCTTGTGTG TAATGGTATG
10901 TTTTATGGGC AGCATATAGA TGGGATTGT TTTAATCTC ATAATCTCTG
10951 CCTTTTAATT GGAATCTTAT GCCATTTATA TTTAATGTGA TTATTGGTCT
11001 CTTTGACTTT AGATCTACCA TGTGCTGTT TGTTCCTGT TTGTCTCATT
11051 TGTTCCTTGT TCCCTTTTCC CCTCCTTCT ACCTCTACTG GATTATTTT
11101 TATGATTCCA TTTTACATCC TTTATTGGCT GATTAGCTGT AACGGCTATA
11151 TTTAGTTTTT GTTTGTTGT TTCTAAGTGA CTGCTTTAAG GTTGTGGTA
11201 AACATCTTTA TCTTGCTACA GTATACATTA AACTTCAACT TCATGTGATA
11251 TAAGAACCTT ATAACAGGCC AGGCGCAGTG GCTCACACCT ATAATCCTAG
11301 CACTTTGGGA GGCTGAGGGA GGCAGATTGT CTGAGCTCAG GAGTCTAAGA
11351 CCAGCCTGGG CAACATGGCG AAACCCCATC TCTACCAAAA ATACAAAAA
11401 TTAGCCAGGC ATGGTGGTGC ATGCCTGTAA TCCCAGCTAC TGGGGAGGCT
11451 GAGGTGGGAG GATTGCTTGA ACCTGGAAGG CAGAGATTGC AGTGAGCCAA
11501 GATCCGTAC TGAAGCTGCA AACTCCTGGG CTCAAGTAAT TCTCCTGCCT
11551 CAGCCTCCTG AGTAGCTTGG TCTACAGGCA TGCAACACCA CATCTGGCTA
11601 ATTTTTTAAA TTTTTTTATA GAAACGGGCT ATGTTGCTCA GGCTGATCAA
11651 AGGTTCTTGG CCTCAAGTGA TGCTCCTGCC TCAGCCAACC AAAGTGCTGA
11701 GATTGCAGAT GTGAGCCACC ATTCCCTGCA GGAACAGTCT TAGATTTATC
11751 CACGTAGTCA CTGTTTCTGG TGCTCTTAAT TCCTTTGTAC AAATCCAGAT
11801 TTCCATCTGG TATAATTTTC CTCTACCTG AAGGATGTTA TTTTTTCTC
11851 TGTTCAGGC CTGTTGGTGA TTAACCTTT CAGCATTTT TTTTTTTTTT
11901 TATTTGACAG GATTCACTCT GTCACCTCAG CTGGAATGCA GTGGTACAAT
11951 TATAGTTTAC TGCAGCCTG AACTCCTAGA CCAAGCAAT CCTCAGCCTC
12001 CTGAGTGGCT GGGACTATAG GCATGCATCA CCACGCCTGG ATAATTTTTT
12051 ATTTTTACTT TTTGTAGAGA TGGTGTCTTC CTACGTTGCC CAGGCTGGTC
12101 TCTTACTCCT GGGCTCAAGG GATCCTCCA CGAGGGATCC TCCCACCTTG
12151 GCTTCCCAA TGTGTTGGAT ACAGGTGTGA GCCACTACTC CTGGTCTTCT
12201 TTCAACTTTT GACTTTTGTA TGTCTGAAAG TCTATTTTGC CTTCATTTTT
12251 CAAAGATAGT TTTGCTGGTT ATAGAATTCT AGACTTTTTT TTTTCTTTCA
12301 GTACTTTTTA AAAAAAAGTT ATTTATTATT TTTTTTATT TGAGACAGTG
12351 TCTCGCTGTG TCACCCAGGC TGGAGTGCAG TGGCACGATC TTGGCTCACT
12401 GCAATCTCTG CCTCCTGGGT TCAAGTGATT CTCTGTCTC AGCCTCCCAG
12451 GTAGATGGAA TTACAGGCGC ATGCTACCAC GCCCGGCTAA TTTTGTACT
12501 TTTAGCAGAG ATGGGGTTTT GCCATGTTGG CCAGGCTGCT TTCGAAGTCC
12551 CTGACCTCAG GGTGATCCGC CCTCCTTGGC CTCCCAAAGT GCTAGGATTA
12601 CAGGCGTGAG CCACTTTGCC TGGCCTCAGG ACTTTTAAAG ATGCTATTCT
12651 GCTGTCTTCT GGCCCACTTC GTTCAAGAA GTGTGCTGGC ATTCTTTGGT
12701 CCTCTGTATG TAACCTTTTT ATCCACCGCT TTTAAATAT TCTTTATCAC
12751 TGGTTTTAGG CAGTTTCATA ATGTGGCTTG ATGTAGTGT TTTTGTGTTG
12801 TTTGTTTTTT GTTTCTGTG TTTATGCTTG TGATTCATTG AACTTCTTAG
12851 ACATGTGGGT TTGCAGTTTC ATCAAATTG GAAAAATAAC ACCCATCTTT
12901 TCGTGAAATC TTTGTACTGT CCCCTATGC GATTCCACAG ACTTCTTTG
12951 CAGAATCCCC TGTCTGTATT TCAGGCACCT CGAAGGTGTT CTCACAGCTC
13001 ACTTGGGATT TTCAGCCTTC TCCTTTGTGT GTTTCATTTT GGATAGTTTT
13051 TGCTGCTACA TTTTAGTCTT TTTTCTCTGT AATGTCTCAG CTGTTGTAAT
13101 TGCACCAAGT GCTTGTTCCT TCTCAGACAC TGTGGTTTTT ATTTCTAGAA
13151 ATCCAGTTTG AATCTTTTTT TTTCTTTTTT TTTTTTTTTA AGATGGAGTT

13201 TCGCTCTGTT GTCCAGGCTG GAGTGCAATG GCGCGGTCTT GGCTCACTGT
13251 AACCTCTGCC TCCCGGCTTC AAGTGGTTTT CCTGCCTCAG CCTCCCAGGT
13301 AGCTGGGATT ATAGGCATCC ACCACCATGC CCAGCTAATT TTTGTATTTT
13351 TAGTAAAGAC AGAGCTTCAC TGTGTAGGCT GGGCCAGTCT CGAATTCCTG
13401 ACTTCAGGTG ATTCAGCTGC CTCGGCCTCC CAAAGTGCTG GGATTACAGG
13451 CGTGAGCTAC CGTGCCCAGC CTCAATTTGA ATCTTTTAAA ATGTATTTTC
13501 TACATCTTTA CTTAACAGTT GCAATCTTTC CTCTGTTTTT TTTTAAAGTA
13551 TATGGAATTT GGTACAGTT ATTCTTTTAA TGTCTATTA ACTAAATTAC
13601 ATTTGGTTAC AGTTACTCTT TTAATGTCCT GTTAATTCTG TCATCTTTAA
13651 CTATGTTACT TTTGGGCTT TTTTTATTT TTTATTTTTA TTTTTGAGA
13701 CAGGGTCTCA CTGTGTCACC CAGGCTAGAG TACAGTGGCA TGATCATGGT
13751 TCACATATAG CTCAACCTCC CAGGCTCAAG TTATCATCTC ACCTCAGCCC
13801 CCCAAGTAGC TGGGACTATA GGCATATGCC ACCATGCTCA ACTAATTTTT
13851 TATTTTTTGT GGAGACAGTG TCTCACTAAT TTGCTCAGGC TGGCCTCAAA
13901 CTTCTGGGCT CAAGCAGTCC TCCTGTCTTG GCTTCCCAA GTGTGGGAT
13951 TACAGGATG GGCCACTGCA CCCAGCCATA TTTTCTGCT TCTTTGCATG
14001 CCTAGTAATT TTTGTTTGA TGGATGCTAG ACTTTGTGAG TTTGAGGTTA
14051 TTGCGTGCTG GATATTTTTG TATTCTATAA ATCTTCTTGA GCTTTGTTCT
14101 GGGATACAGT TAAGTTACTT AGGACAGTTC GATCCATTTA GGTCTTACTT
14151 TTAAGCTTTG TGAGCTGGAC CAGAGCAGTG TTCATTCTAG GGCAAGCCTA
14201 ATTTGGCCCA CCACTTGATT TTGTAATAA AGTTTTATTG GAAAATCACT
14251 CCCATTCAAT TACATGGTCA ATGTCAGGTT TTATGCAATA ATAACAGACT
14301 TGAGTAGTTG CGCCAGAGAC TTTATGGCCC ACAAAGCAAA AAGTTTTTAG
14351 TATTTAGCCC TTTACAGAAA ACATTTGCCA ACCCCAGGCC TTGGGCTAAT
14401 CTTCCCCACT GCTGAGGTAA AACCTCTGG AGTCCCTGAT GCTCCATGAA
14451 TTTATGAGGT TTTTACTCT GGTGTGAATC CTGGGATTG TCCCTTGTA
14501 CCCTCTTTGT GACTCTTCCC CACTGTGGGT TTGGGTCATT TTCGCCAGC
14551 TTTGCCTCGA CCAGTTCTCT GCTGCACATC CGAGGGCACC TCTGCAGATC
14601 TCTGCAGCTC TGTCTCCGGC TGCCCTCTCT GCTCCTGTGA CCTTGGGGTC
14651 AGCTCCAGCT GCAGGCCTCC TCAGACTTCC TCCCAGCATT GCCCCCTGAA
14701 CTCAGGGAAA CCCCTGGGCT CTGCTAGGGG GTCCCTCCA TACACATGGG
14751 GACCATCGGG GGAACCGTA GGCAGCGGC GGGGCAGTGA TAGACTCACC
14801 TCCCTTGTTT CCCTGTGGCC GTCATGGCTG CTGTCCCGCG TCTTAGTGGC
14851 CATCATGTCC ACTGTCCCGC ATCTTGGAGG CCATTGTTT TGTGTGTTGG
14901 GAAAGGGGGT AAATCTGGTT TCTTCTCTCT TTGTGCTTTG AGGTGGATGT
14951 GCCTGGGGCC ACCTGATTTC AGAGAGTCTT TGCCGCGGTG CACGATGTCC
15001 AGTCACCTGG AGCTTGGCAG CAGGTGGCAT GTGCACCTGT CCGCAGCCTG
15051 CAGCTCTGCC GTCCACCTG CTTATGCTGC CACATAGCAT TTATGTCTGT
15101 GTTGTGTTTC CCAGAGTGCA GAGTGAGGAC TGTGGCCCGG TGGGGGCGTC
15151 TCCTGCCTCT GGGATGCTG CCCCAGTAAC CTGGAGGCAG CCGCGCTACC
15201 CCACACTCGC CGGGAGCAGG GTCTCGGACG CACCTCTCT TCTCTCTAG
15251 CTCGCAACGC CCTGGTGGTC TCCTTCGCAG CCCTGGTTGC GTACTCCTTC
15301 GAGGTGACTG GATACCAGCC TTTTCATCCTA ACAGGGGAGA CAGCTGAGGG
15351 GCTCCCTCCA GTCCGGATCC CGCCCTTCTC AGTGACCACA GCCAACGGGA
15401 CGATCTCCTT CACCGAGATG GTGCAGGTGG GCGGAGCCGG GAGGCAGGAT
15451 GGCGTGGCTG AGGCTGCAGT GGCCCTGGC CTGGCTCCTA CCCTGATGTA
15501 TCTGCTGGGT GCCAGGGGGT CTGAGGTCAG TTAGGACAGC TGAGTCTCA
15551 GGAACGGACA TCTCAGTTAT TAAAGAATCC CAGGTGGAT GCAAACCTCAG
15601 CGAGCTCAGG GATGTCACGT TTGTGTTTCT GGGCGCTTCT CCTGTTTTGG
15651 ACTCCAGCTG AGGATGAATT TACCGTGTTC CTCCAGCAC CTGGCGCCTC
15701 TTCAGACAAG GAGGCGGATC CTGCAGCTGA CAAGCACTTG CTCCTGTTAC
15751 CTGTGGGGCG GGGTGGGTCC TTGCTGCTTT CATGGGTCAC TGCTGGGTCC
15801 TACCCCTTAG GAAGGTCACT CACCATCCCT CTCTCCTCTC TCAGGACATG
15851 GGAGCCGGGC TGGCCGTGGT GCCCTGATG GGCCTCCTGG AGAGCATTGC
15901 GGTGGCCAAA GCCTTCGGTA AGACGCTGT CACCCACACC CCAGGTCTCC
15951 CAGTGCGCCG GCTGGGCTAG GCCTGCCTGC TTTCTAGCTT GCCTTTATCC
16001 GTTACTAGTT TTAGAAATTT GAATTCATAT CCAAGTAATA CATGCTCATG
16051 ATAGATACAT ATGTATTGTG TATATATGAT AAACTGGAT CTATAATGAG
16101 GCATGCCCTC CCACCCCATG GTGTGCTGGT GAGTGTGTA ACAGCCTCTG
16151 CTGTTTGTGG AAATAAAGG TTTTGCTTCG TGGCCCTTGC CGATGTCCAC
16201 GGTGTAAACG CTGCTGTCTG ATTTTAAGGT AACGTCACTG AAAGGGGAGT
16251 TTGCACATGG AGCTGGGTTG AGATCTGCAT GAACAATCAT ATTCTATGGT
16301 GTCTCCACCA TGATAGATACA GTGGGTGCAA ATAACCTCAT CAGTAGTAGC
16351 CAAATGCCAA ATAAATTAGG AAGTGATGAG TTTTAAAGTAT TATCTTTGGG
16401 CCAGGCATGG TGGCCAGGC CTGTAATCCC AACACTTTGG GAGGCTGAGA
16451 CGGGAGGATC GCTTGAGCTC AGGAGTTTGA AACCCACCTA GGCAACGTAA

16501 CGAGACCTCG TTTCTATTAA AAATAAAAAAT TAGCTGGGCA TGGAGCACAC
16551 CTGTGGTACC AGCTTCTCAG AAGGCTGAGG CAGGAGGATC ACTTGAGCCC
16601 GGGAGGTCGA GGTGGCATTG AGCTGTGATA GTGCCACTGC ACTCCAGCCT
16651 GGGCAGCAGA ACAAGATCCT GTATTATCTG TTTAATTGAA AGTTTAAAT
16701 TAATTGGTAA TAATGGCTAT GTTAGTAAT AGGCTCACAA AACTCCTAAA
16751 CATTCAGCAA CATGCTTAT TCAGCTGGCT CAGCCCATCG GCCAGCCCCCT
16801 CCTCGCCCCC AGGGAGGCAG GGAACCCCTC GTCTCCTTCA GTGTCTGTTT
16851 TGAGTCAGCA TCTCTAGATT CCCTCCTTGC AGCTCCGTCC TTCACTCGCT
16901 CCACTCCTGC CAGCTTTGTA TTGTCTGTGC GGATGACACC TTCGTTCTGT
16951 CTTGTACCCG TCATTTAGTC CTCTTGCGT TTGGGTTTAT GTTGATTCTA
17001 AAGGTTACAG CCCAGTTACC AGTGTATGTT ACTGTGACTG TGTAATACGT
17051 GTTTACTGCC TAGCCCAGGC TGTGCCAGG TAAAGTTCCCT TCTCCAAAGT
17101 CCCAGTGCTC TAACCCTTAC CCCCTTCAAA GGAGTTTTTTT TTGTTTTTTT
17151 TTTTTTTTTT TTTTGGAGA CAGAGTCTCA CTCTGTGGCC CAGGCTGGAG
17201 TGCAGTGGCA CGATCTCTGC TCACTGCAAC CTCGCTTCC CAGGTTAAG
17251 CAATTCTCAT GCCTCAGTCT CCCAGGTAGC TGGGAATACA GGCATGCGCC
17301 ACCACGCCCA ACTAATATTC TATACTTTTA GTAGAGACAG GGTCTCACTA
17351 TGTGGCCAG GCTGGTCTCA AGTGATCTGC CCACCTGGGC CTCCCAGTAA
17401 TTTTTTTTTT TTTTTTTTTT GAGATAGTGT TTCACTCTTG TTGCCCATGC
17451 TGGGTGCAAT GGTGTGATCT TGGCTCACCG CAACCTCCGC CCCCAGGTT
17501 CCAGCAATTA TCCTGCCTCA GCCTCCCAAG TAGCTGGGAT TACGGGTGTG
17551 CACCAACATG CCTGGCTAAT TTTGTATTTT TAGTAGAGAT GGGGTTTCTC
17601 CATGTTGATC AGACTGGTCT CAAACTCCTG ACCTCAGGTA ATCCGCCAC
17651 CTCAGCCTCC CAAAGTGCTG GGATTACAGG TGTGAGCCAC CATGCTCAGC
17701 CCTCCAGTAA TTTCTTTTTT TTTTTTTTTT GAGACGGAGT CTCGCTCTGT
17751 CGCCCAGGCT GGAGTGCACT GGCGTGATCT CGGCTCACTG CAAGCTCCGC
17801 CTCTGGGTT CACGCCATTC TCCTGCCTCA ACCTCCCGAG TAGCTGGGAC
17851 TACAGGCGTC CGCCACCACG CCCAGCTAAT TTTTGGATT TTTAGTAAAA
17901 ACGGGGGGTT CACCGGGGGT CTCGATCTCC TGACCTCGGG ATCTGCCCTG
17951 CCTTAACCTC TCAAAGGGCT GGGATTACAG GCGGGAGCCC CTGGGCCCCG
18001 GCCGTAATT TTTAATGGAA AAACAAGGGC TCACTTTGGG GGCCAAGGCT
18051 GACCTCAAAC TCCTGAGTTC AAGGGATCCT CCTGCCTCGG GTTTNNNNNN
18101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNTGGTCT
18151 GGGCTAATTT TTGTTTTTTT GTAGAGAGAG TATTTGCCAT GTTGCCAGG
18201 CTGGTCTCGA ACTCCTGGGC TCAAGCGATC CTCCTGCCTT GGCCTCTAAA
18251 AGTGCTGCGA TTATAGCTGT GAGCCACCAT ACCTGGCCTG ATTACATGGT
18301 ACTGTACCTC TTTCCACTCT GTGATACATT CAAGCACATG TTTACACGCA
18351 CATATATGTG AATATTTTAG GTGTGTATTC TTAGATTAC AAAACTTATT
18401 GCCGTGTGTA GAAATTCATA ATCTGTGTTA CATTTTGTGA ATACTTTTGT
18451 AGCGATGTTG AATATTACTG ACCAGGTCTT TACCGTTGGT TCCCTGTAGC
18501 ATCTCAGAAT AATTACCGCA TCGATGCCAA CCAGGAGCTG CTGGCCATCG
18551 GTAAGACCCC AGCCGCGGGA AGGAAGACAC CAGCTGTGGG CCTCCAGGGT
18601 CCCAGGCCTG CCCCTCTGTG TCTCCTGCAT TGTAGGAAGA CCATGATGGT
18651 GGTGATGAAC TGGGAGGGCA AAGGTGGCCC CAGATGGGAT CTCTGGAAT
18701 ATTTAGTTTT GTTTTGGGTT TTTGAGATGG AGTCTCACTC TGTGCCCAG
18751 GCAGGAGTGC AGTGGCACGA TCTCGGCTCA CTACAACCTC TGCCTCCTGG
18801 GTTCAAGTGA TTCTCCTGCC TCAGCCTCCC AAGTAGCTGG GATTAAAGGC
18851 ATATGCCACC ATGCATGGCT AATTTTGTGA TTTTGAGTAG AGATGGGGTT
18901 TCACCATGTT GGCCAGGCTG GTCTCGAACT CTTGACCTCA AGTGATCCGC
18951 CCGCCTCAGC CTCCCAAAGT GCTGAGATTA TAGGCGTGAG CCACTGCGCC
19001 TGGCCTATGC CTTGTATCT TAAACCTTGA GACTCAGAGT GTGGCCTGGG
19051 ACCAGCAGCC TGGGCACTGG GCATCTCCTG GGAGCTTACT GGAAAGGCA
19101 GATCTCAGGC CCCACCCAG ATCTCCGAAT CAGGATCTGT ATTCTTCAGG
19151 GCACGCCCAG GGGATTCATG GGGTCAGCTT AGAAGTCAA GTTTGGAAGC
19201 CGGGCATGGT GGCTCACGCC TGTAATCCCA GCACTTTGGG AGGCCGAGGG
19251 AGGCAGATCA CCTGAGGTCA GGAGTTTGAA ACCAGCTTGG CCAAACCCCG
19301 TCTCTACTAA AAATACAAAA ATTAGCTGGG TGTGGTGGTG TGTGCCTGTG
19351 ATCCAGCTA CTTGGGAGGT TGAGGCAGGA GAATCACTTG AACC CGGTA
19401 GCGGAGGTT ATAGTGAGCC AAGATTGCAC CACTGCACTC CAGCCTGGCA
19451 ACAGAGCGAG ACTCCTTCTC AAAAAAAAAA AAAAAAAGT CAAAGTTTGG
19501 GAAATGCTTA GAGACCCCAT GTTTTTCAA GACTTGTITT STATGACTTT
19551 ATGAGATGAT CAAGCAGGTT TGGCCACAAG GGGAGACAGC CAAAGGCTCA
19601 GGAAGATAAC CAGTTTTTTT TACTCAAAG GTCCCTTAGT TACCAGGGAG
19651 GCCACAGCAT GCCTACTTAG GCCCATGTCA GAGGACACCA GGGGTGTCCC
19701 CGGAAGGATA AATCCCAATC TGTGATTTTA GGCTCCGCC CTTTTCGGG
19751 AATTTCTTTT GCTTAATGCT AATTTCAATA AAGGCCCCCG GCTTAAGACT

```

19801 TTTTCATATGT TATTAATTTT AGGCGGGTTC TAACTCCTTA ATGAAGAAAT
19851 ATTCTCCCTT CCCTGGATTT CCGAATCACA CCATTTCTTC GAGAAACGCC
19901 TCCCACNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19951 NNNNNNGTTT TTTTTTTGGG CACCCGGCCA TCATCTGGGG AGGCAGCATT
20001 AAGGCCCATC ACTTTGCCCT TCAATGTGAC TGTCTCTTTT TTACACGAAA
20051 AAGTTTTTTG GTGGAACCC GGTGGTCGGC CCCATTACCT CCTGAGTGCC
20101 TAGGGTTTGT GTTCCGGGCC ATTTATCTCA GCTTTCACCC CCTCCTGGAT
20151 GGTAAAGATGC TCCCCTCTGT TCTGCGTGGG GAAGGCGGGT CCTGGTCCCA
20201 CAGGATAGAT TGGGTCAAAA GTGTTTCAGA GCTAGGATCT CTTCAGGCT
20251 CTCGTGGGTC ACTCCTGTTT CCCAGCACTT TGGGAGGGTA AGCCGAATGG
20301 GTCACCTAAG GTCAGGCATT TGAGACCAGA TTGGCCAGCA TGGCACACTC
20351 CCTGTGTGCT ACTAAAAATA GAAAAATTAC ACTGGGTGTG GTGGCACATG
20401 CCCGTAGTCC CAGCACTTTG GCAGGTTCAG CCGGGTGGAT CACTGAGGTC
20451 AGGAGTTTGA GACCAGCCTG GCCAACATGG CAAACTCTG TCTGTACTAA
20501 AAATACAGAA ATTAGCCAGG CGTGGTAGTG TGTGCCTGTA ATCCCAGCTA
20551 CTCAGGAGGC TGAGGCAGGA GAATCGCTTG AACTGGGAGG CGGAGGCTGC
20601 AGTGAGCTGA GATCGGCCCA CTGCATTCCA GCCTGGGCAA CAAAGCCAGA
20651 CTCTGTCTCA AAAAAAAAAA AAAAAAAAAA AAAGGCTAGG CTTTACGTCT
20701 GCAAGAATGT GGCCTGTTTT CTTCCTTCCC TGAAGGAGTG CGTAGGCCCA
20751 GGCCCCCAGG ATGATTCTCC CGAGCCCTGC CTTCCTGCCT CCCTTCTGGG
20801 TTTTGTGCTC CCTCCCCATC TCCTTCCCT CCATCCTGTG TGCCTTCCCT
20851 CCACGATCAG CCTGTCTTGC CTCCTCCCCA GGTCTACCA ACATGTGGG
20901 CTCCCTCGTC TCCTCCTACC CGGTCACAGG CAGCTTTGGA CGGTGAGTGA
20951 CATGTCCGCC TCTTCTGTTT GCCCACGTTG GACGCTTAA CGTTGTACG
21001 CTGACAAGGA GTCTGCCTGC CCTGACCCCG GCGCCCCGTC CTCCACTGTG
21051 AACGCTCCGT GGAGAGGCAG GGCTGGGGGT CACCCACTGT CCTCCAGGGT
21101 GTTCTCTGTT TCTTTATTCT CATAGATCGT CCTGCAGTTT CATACTAGAA
21151 AGTTCCACTG GGCATTGTGG TACACCCTTG TTATCCATT TACTCTGGAG
21201 GCGGAAGCAG GAGAATCGCT TGAGCCCAGC AGGATGAGGC TGCAGTGAAT
21251 TGTGACCGCA CCGCTGCACT CCAGCCTGGG CAACAGAGCG AGACCCTGTC
21301 TCTAAAAATA ATACTAGAAA GTTCCCAGCA CGCCAAAGCC CTCCTAGCTC
21351 CTGGTGCCAG AGTCAATTC TGAAGGACG TGGAGATAGG AAGGGCCTCG
21401 GCTCTGTCTT GAAGCAGCCG GGCATGAAGC TTAGCCAGA TGCCCTACGG
21451 CCCCTCTCA GTCCAGACAA CAGGATGGAG GTGACCTGTG GCTTAAAGGA
21501 GAGAAGGAGG CGTCGCCTGG CACTGCCCAG TCCCCAGCT GGTGACCCTT
21551 GCCCTGCTGG GTATGGGGGC CCCACCTGGA TGGGGGAGG AGACAGAGTC
21601 GGCAGGAACC TGAAAGGACA CGTGCTTCCT GAGCTTCTTC CTATAGTCAG
21651 GGTGGCCCAA GCGCGGCTGT CTGTGACTGC ACCCTAAGTC TCTTTGCCTC
21701 GGTCCCCCTG CAGTCCCCGC CTGCTTCCA AGCCGTGCTG GGAGCTGACG
21751 TCCCCTCGGA AGATCAGCCA CAGGAGTGTG GACTGAGGTC TCCCTTTTCC
21801 CGGCCCTG TGACTGACGG TCTCTGTGTT GCCTTCCAGG ACAGCCGTGA
21851 ACGCTCAGTC GGGGGTGTGC ACCCCGGCGG GGGGCCTGGT GACGGGTAAG
21901 GCCCCCATC TTCCCCTGT GCGCGCAGCC CTGAGAGTGG GAGAAAGGGA
21951 GGAGGGGGCC CACAGAGACG TCCCTTTGGC TCATGGGCGG TGCGCCCCGG
22001 GACTGCACAG GGACTTGGGG GGCCACACAG GAGTAGGGGG ACCACAGGAG
22051 ACTGAGCAGG GGCTGGGGGC CTTGGCAGTC GTCGCCCTAC CCCCACCCCT
22101 GTCCCCAGTG GGCTCTGCTG AACAAGAGGC TGCTACGCTG CGTGCTGGGG
22151 GGACCTGCA CTCCCAGGT CACCTGTGTT CCGGTGCCCC GCAGGAGTGC
22201 TGGTGTGCT GTCTCTGGAC TACCTGACCT CACTGTTCTA CTACATCCCC
22251 AAGTCTGCCC TGGCTGCCGT CATCATCATG GCCGTGGCCC CGCTGTTTGA
22301 CACCAAGATC TTCAGGACGC TCTGGCGTGT TAAGAGTACG TCCTTGTCTT
22351 ACAGGGGAGA GCGCTGTGAT GCGGTGTCTG AACGCGGAGG GTGTCATTTA
22401 TGCTACCCCA TTTTCTGCA GCCCCTCTG TGGGGCTGGG ACTGGGAAGT
22451 TAGGGCAGTC CCGGAACAGA GAAGTGGATG GCCAGGAGAT GGCCCCAGAG
22501 ATGGTCCCGA GGCTCAGTGG GAAGAGCTGG AGCTCCTTGT CTGACACCT
22551 GGGGTCTTGA GGCGAGCACT GACCCGGGGG AGGGTCCCCCT CCTGATCCCC
22601 CTGCCCCCAT CCTACCCCTC CTTGCCACCC GCCTCCAGCC ACCACTCTGC
22651 CCGGCCCCAG TGGGGGGAGG GACAGGAGAC GTCCCTGGTG ACCAGCAGGG
22701 CCAGCGGAAC AGCCTTGCAC CCTGGCTCAG AATGGCAGTT CCTTTTTTTT
22751 TTATTATTAT TATTTTATT TTTTTTTATT GATCATTTCT GGGTGTCTCT
22801 TGCAGAGGGG GATTTGGCAG GGTCTAGGA CAATATGGGG TTGGGGGTAA
22851 GGTACAGAT AACAGGATCC CAAGGCAGAG GAATTTTCT TAGTGAGAA
22901 CAAAATGAAA AGTCTCCCAT GTCTACTTCT TTCTACACAG ACACAGCAAC
22951 CATCCGATTT CTCAATCTTT TCCCCACCTT TCCCCCTTT CTATTCCACA
23001 AAACCGCCAT TGTCTCATG GCGCGTCTC AGTGAGCTGT TGGGTACACC
23051 TCCCAGACGG GGTGGTGGCC GGGCAGAGGG GCTCCTCACT TCCCAGTAGG

```


23101 GCGCGCCGGG CAGAGGCGCC CCTCACCTCC CGGACGGGAG CAGTTCCTTT
23151 AACTTAACAC ATTTTGT TTTT GTTGTGGAA GAAGCATGGA GTGGGTTTAC
23201 CTAGCTTACT TGGCATTTC TGGTTAGCCC CAGCAAGTTG CCAGGTGAAA
23251 GGATGGATAA TTTCTTGCA TGCCCGTCGC ATGCCAGGGC CTTTCGCGTG
23301 CCAGTGCGCT GGGGCTTTAG AACAGCCCTG AAACGTGTTC ATGTGCCGGC
23351 GTGGAATGGG ATGGCCGCTG CTGCTAGAAA CCAAGGCTCA GCCAGGGTGG
23401 GGCAGGCCTG ACCTATGCGT GCGGTGGAAT CCCCACAGG GCTAAGCCCC
23451 TGCACTTTGT CCCCAGGGCA CCTTCTCCTT GGCCAGGTCT CAAGGGCTCA
23501 CGTGGTCCCT GCCCCACTCC TCAGGCCAGC TCTGTGCCCT GACAAGCCCC
23551 TGCTGCTGCC CTCCTGAGG TTGGAGGCCA GGGACCGGCC GGCAGGTCTC
23601 ACCTGCGCTC AGCTCAGATG GGGAGGGCAT TTCTTTCTTT CGACTTGAAG
23651 CATGGCCTGG TCAGCAGCTG CTGTCCCAA GTCCCTCAGG GCTGCTTGGG
23701 GTCATGAGC ACCTTTACTC ATATGTGGGG GGCAGAAGGC TGTCCCGCTG
23751 CTCAGCAGGG CCATGTTGGG GCCTCGGGCA GCTGCCGGGC ATTCTCAGC
23801 TGTGCCCTTC TCCTAGGGCT GGACCTGCTG CCCCTGTGCG TGACCTTCCT
23851 GCTGTGCTTC TGGGAGGTGC AGTACGGCAT CCTGGCCGGG GCCCTGGTGT
23901 CTCTGCTCAT GCTCCTGCAC TCTGCAGCCA GGCCTGAGAC CAAGGTACCC
23951 CTCCGTGGCC TCTGAGTGGG GAGTGTGCTG GGGGCAGGAT TCCTGGGCAT
24001 GGTCTTATGT TTTGAGGGTC CGGGGTGATT GTGGTCTGTT GTGCTGCTGA
24051 AGGGGACCGC TCGTGGCAG GTGGGCAGTC ACCTTGCTAT AAACCATGGT
24101 GTTCTCCAC TGTGTGGGG CCCTGGGGGT CTCCCTTAG CACCCCTCTC
24151 CCGGTCCCTT GCAGCACGCT AGGTGGGGT GGGGCTTCCC GCTTGGGACA
24201 GGCCAAGCCT GGTGGAGGCC ACCCGGTGAG ACCCGCTCC AGGACTCACT
24251 CCTCCCCACA GGTGTCAGAG GGGCCGGTTC TGGTCTGCA GCCGGCCAGC
24301 GGCCTGTCTT TCCCTGCCAT GGAGGCTCTG CGGGAGGAGA TCCTAAGCCG
24351 GGCCCTGGAA GGTGCATGGG CGGGGGTCAA GGTGGTCTGA GGTCACTCCC
24401 CTGTCTCTG CCCCCACTC CCTGCTGTTT AGGACCCCAA GACCCTGTCC
24451 CCGACGCTCT CCAGTCCACA AGGATGCAGG CATCTCTGAG TGGGCTGGAC
24501 CGTCTCTGT GGGCTCAGC CAGTGGCTGC TGCAGCAAGG GTGGTGGCTC
24551 CCCACATATC ACTCCTTCCC TGCCCTAAA GTCCGGTTCC TGTTTCTGGG
24601 GGGTTGATTT TAGGGGAGCT AAGGGCCTGT GAGTCTTAGG AGGGAACAG
24651 CTGCTGCTGT CACCAAACAA TTGTCTCTGG TCCTGCCACC CGAATCCCC
24701 AACTGGGCGA CTCAGCCGCC ACGAGATGGA GCACTCTGGC CTCTCTGTCC
24751 CCTGCCCCTG GCCAGAGCCT CCTTGGCCT CTGCAGAGCA GCTTTGGGCT
24801 GCTCTGGGTG GCGTGACCTG GCTCGGGCCT GTCTCCCCAG TGTCCCCGCC
24851 ACGCTGCCCTG GTCCTGGAGT GCACCATGT CTGCAGCATC GACTACACTG
24901 TGGTCTGGG ACTCGGCGAG CTCCTCCAGG ACTTCCAGAA GCAGGGCGTC
24951 GCCCTGGCCT TTGTGGGCCT GCAGGTGGGT GTGCACTGGG ATGCCTTAGG
25001 GGTTAGCAGC TGCCGGAAGG CCTTCTGTG CCTGCCCTCC ATGGCGAATG
25051 TGACATCTCT GGGCTGTGAT GCTGGACGGC CCTTCGGCCG GTGCTGGCTC
25101 TGCTTCTGAT TTAAACAGTT CTTGTCCCCA TCTGGCCTTC CTCGTCCCTC
25151 CCTGTGGAAG GGGGAGCGGT GGCCCCCAGC CCTCCGAGGG GTCACGTTAT
25201 GGCTTCTGGT CACTGCCACA CTGTCTTTG TGCTGGGGAC ACACAGTGAA
25251 CGAGGGTCAG TCCCTGCTCT CAGGGGAATT GTATTTTAGG AAACAAAAG
25301 AGACAGGTGA GGCTAGGCGT GGTGGCACAC TCCTGGAATC CCAGCACTTT
25351 GGGAGGCCAA GCGGGGCAGA TCACCTGAGG TCAGGAGTTC AAGACCAGCC
25401 TGGCCAACAT GGTGAAACCC CGTCTCTACT AAAAATACAA AAAAACTAG
25451 TCAGGTGTGG TGGCGAGCGC CTGTAATTCC AGCTACATGG GAGGCTGGGA
25501 GAATCGCTTG AACCTGGGAG GTGGAGGTTG CAGTAAGCCG AGATCCCACT
25551 ACTGTACTCC AGCCTGGGCG ACAGAGTGGG ACTCCATCTC AAAAAACAAAC
25601 AAACGAAAAA AAACAGGTGC TCATAGAATT TCATGAAAAA CGTATTGTCA
25651 GGGCTTCCAG AGGCTGAAGA CGGGTTTCTA TGGAGGCCGT CCTGTTTCTA
25701 GCCGAGGTA AAGTGAAGG GCTGGGTCCC AGGCCCTGCG TCTTAGGCCT
25751 CACCTAGGAG CTTTCTGAGC ACTGCAGGGT CAACATCCCA GGGGTGTGGC
25801 CAGTGTTTGC AGAGAGGCAG GGGTCTCTGT TGCTGTGGTT AAATGTGCGC
25851 TCTCTGCCAA GTCCTAGATG GCAGAACGTG GGGGACTAGG GCGTGTCCCC
25901 AGGCGCCCGAG AGGAGACAT CATGAACATG CCATGGAACA GAGAGCCAG
25951 TGACCCGTAT ACCCCAGGTG TGGACCACAG CCGACCCCTG TCAGAGTTTC
26001 CTTTCTTCCC AACCTGGCCC ACGAGGCTAG TGTATCTTC ATCCCCATGT
26051 CCAGAAGAGG TCACTGAGGC CCAGAGAAGC CAGGTAATCT GCCCACGGTC
26101 ACACAGGATG GTGGGCCAG CTTTCAATTT GGGCATGTGC CTTTCAAGCT
26151 TCGTCTATAA GTGTTACGTC CTGTGTCATA TTTGCAGAGC ACATGTCTTC
26201 ATGGTCTTTG GAGATGATT ACTTAAAAA AACTCTCTT GACCCAGACA
26251 TGGTGGCTCA CACCTGTAAT CCCAGCACTT TGAGAGGCCG AGGCGGGAGG
26301 TCAAGGGTTT GAGACCAGCC TGGCCAACAT AGTGAACTC CGTCTCTACT
26351 AAAAAACAA AAATTAGACG GGCATGGTGG CGCACGCCCTG TAATCCAGC

26401 TGCTCAGGGG GCTGAGGCAG GAGAATCGCT TGAACCCGGG AGGCAGAGGT
 26451 TGCACTGAGC CGAGATCGCG CCATTGCACT CCAGCCTGGG TAACAAGAGT
 26501 GAAACTCCAT CTCAAAAACA AACAAACAAA CAAACAAAAA GTCCTCTGGG
 26551 CCAGGCGCTG TGGCTCACAC CTGTAATCCC AGCACTTTGG GAGGCCGAGG
 26601 TAGGAGGATC ATTTGAGGCC AGGAGTTTGA GACCAGCCTG GGCAACATGA
 26651 CGAAACCCCA TCAAAAAAAT AGGAAAAAAT AGCTGGACGT GGTGGTGTGC
 26701 ACCTGTAGCT ACTCAGGAGG CTGAGGTGGG AGGATCACTT GAGTCTGGGA
 26751 GGTTGAGGCT GCAGTGAGCC ATGATCGTGC CACTGCACTC CAGCCTGAGC
 26801 AATGAGCAAG ACCCTGTCTC AAAAAACAAA AAATTAAAAA AAAAAAATC
 26851 TTCTCTGAC AGCATTCCCC TGGGGCTGCG TTTCTTCTCA CCATTCACTG
 26901 GTATGGAGGT GAAGCCATAC CTCTCCGGGA GACTCTGAGA TGGCATGTCT
 26951 ACCAGGCTGC CGACCCGTGT GCTACAGAGG AACATCCCTG CCCTGGCTAA
 27001 AGTCTGTCTG TCTCTCAGGT CCCCGTTCTC CGTGTCTGCTG TGTCCGCTGA
 27051 CCTGAAGGGG TTCCAGTACT TCTCTACCCT GGAAGAAGCA GGTGGGCACA
 27101 GTCAGACATC CTGTGGCTTT GGTGATTTTG TAAAAATCAT AAATGCTTAT
 27151 TGTAAAAAAT ATGGGAAACC AGCTGGGCAC AGTGCTCAT GCCTGTAATC
 27201 CCAGCACTTT GGGAGGCCGA GGCAGGTGGA TCACCTGAGG TCAGGAGTTC
 27251 GAGACCAGCC TGGCCAACAT GGTAAACCT CATCACTACT AAAAATACAA
 27301 AAATTAGCTG GGTGTGGTGG CATGCACCTC TAGTCCAGC TATTCCGGAG
 27351 GCTGAAGCAG CAGAATCGCT TGAACCCAGG AGGCAGAGGT TCACTGAGC
 27401 CGAGATTGTG CCTCTGCACT TCCAGAGTAG GGTGAGAGAG CTAGCCATGT
 27451 TGTCCAGAAA ANNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27501 NNNNNNNNNN NCCTGGTTGA GCAGGTATAA TCGCTTGCAA CCCAGGTAGC
 27551 AGCAGTTGTA CTGAGTCCGA GAATTGTGCC ACTGCACTCC AGCCTGGGTG
 27601 AGAGCAGCGA GACTCTGTC TCAAAAAAAA AAAAAACAAA AAAAAACAGG
 27651 AACCACCTCC ACCCCAGTCC ACTTTGGTGA TCACTCCATA CCCCTCCCTC
 27701 AAACACGCAT ATGTACCTGC CTGTGAGGAT GTGGATATAT GTTTGCGGTT
 27751 TTACGTAAAT GGGACCAATT CATACTGGT GCTCTGGAAC CCACATTTTT
 27801 CATGCAGAAG GTTGGAAGGA TGTCTTCCA GCCGAAAGTC CACATCCCGG
 27851 GGGATCAGGA CAGAGCAGGG CCGGGTCAGG AGATCCAGAA GCCCCGGGAC
 27901 AGAAGGTACG GGAGGGACAG GAGCAGGGTG GCGCTGACC CTTGAGACAG
 27951 CAATCGCAGG AGGTCTGAGC CGCAGCAGGT GTCAACAAGA GGATGGGCCA
 28001 GAGATGCAGA GCATCCACCC CAGGCCACAC AGCAGTGGCC AGAGGGTCCC
 28051 AGGCCCCAGT GCTAGGCCTC TTCCTCTTCC ACTGAGGTCA CAGCTGAAGC
 28101 TGGGTCAGCT CCGTGAGAGT GAGGGGTGGC GGATGTTGTA CTGACTTCCT
 28151 TGGCTCAATG TGACGTCAGG GAGATTCACC CATGTTGTTG TAGAATCAGC
 28201 TCAGATGCAG TGCACTGGAG GTGGATAAGC AGAATGTGGC CTGGCCGTGG
 28251 GATGGGAGGG TACCCTGGAG CAGTAAGAAA GGGCCGTTAG TCACCTGAAA
 28301 AATACGCTTA CAGAGACTCA GGTGAGACCC TCATGGAGTT AGTGACACTG
 28351 GCCTGGGTGG CCCACAGCTC CTTCTGTCAC ACCTTCCAGG ACTCTGGAAG
 28401 GCCCTCCTTA ATCCCTTCCT GTGAACTGAC CCATCCTCAC TTCTGAGCTT
 28451 TTAGTGCTTG AAACATTTAT TGTATTTTCT GCAGAGAAGC ACCTGAGGCA
 28501 GGAGCCAGGG ACCCAGCCCT ACAACATCAG AGAAGACTCC ATTCTGGACC
 28551 AAAAGGTTGC CCTGCTCAAG GCATAATGGG GCCACCCGTG GGCATCCACA
 28601 GTTTCAGGGG TGTTCCGGAA GGTTCTTGTC ACTGTGATTG GATGCTGGAT
 28651 CCCGCTGAT AGACATGCTG GCCTGGCTGA GAAACCCCTG AGCAGGTAAAC
 28701 CCAGGGAAGA GAAGGAAGCC AGGCCTGGAG GTCCACGGCA GTGGGGAGTG
 28751 GGGTCACTG GCTTCTGTG GGATGACTGG AAAATGACCT CGCTGCTGTT
 28801 CCTTGGCATG ACCCTCTTTG GAAGAGTGGT TTGGAGAGAG CCTTCTAGAA
 28851 TGACAGACTG TGCGAGGAAG CAGGGGCAGG GGTTCACAGC CCGGGCTGTG
 28901 CGAGGCATCC TGGGGCTGGC AGCACCTTCC CGGCTACCA GTGCCACCTG
 28951 CGGGGGAGGG ACGGGGCAGG CAGGAGTCTG GGAGGCGGGT CCGCTCCTCT
 29001 TGCTGCGGC ATCTGTGCTC TCCGAGAGAA AACCAAGGTG TGTCAAATGA
 29051 CGTCAAGTCT CTATTTAAAA ATAATTTTGT GTTTTCTAAA TGGAAAAAGT
 29101 GATAGCTTTG GTGATTTTGT AAAAGTCATA AATGCTTATT GTAAAAATA
 29151 CAGGAAACCA CCCCTCACCC TGTCCACTTG GGTGATCATT CCAGACCCCT
 29201 CCCCAAACAT GCATATGTAC CTGTCCGTCA GTGTGTGGAT GTATGTTTAC
 29251 AGTTCTACAT AAATGGGATC ATTTTATACA TGGTGTCTG GAACCCACAT
 29301 TTTTCATGCA GTCATTTGCA GTGAATTATT TATTGTGATA ATAAATAGCA
 29351 TTAGAATACA AGATTTTAA TGTCTGCGTG GTATTTTGGT CTATATATGC
 29401 ATCATAATTG ACTTACCGAG CCCTCTGTTT AACGTGTGCG TGGGTTAGAG
 29451 ACGGGATCGT GCCTCCTTTA GATGTGTCAG TTGAAGCCCT TGCCCTCTAT
 29501 GTGACTGCGT TTGGAAACAG GGCTTTTAGG GGGTAACTAG GTTTAATGAA
 29551 GGTCATAAAG GTGGGCCCTG AACCGATAGC TTTAGTGTCC TCGTGAGAAG
 29601 AGACGCCAGA GAGCTCGTTC TCTGCACCT CACACCCAG GAACTCCAT
 29651 GAGAGGACAC GGCAAAACCA GGCCATGCGC CCACCAGGAA GAAAGGCCCTC

```

29701 ATGAGGACCC CAGCCTCCAA AACTGGGAGA AGATGAATCT CTGTGCTAGG
29751 CCCCGCAGCC TGGGGTGATC TGTGACGGCG GCCTGAGCAG GTGAGGACTG
29801 CCTGCATGTT TGTATTTTAA TGAATGCTTT GATTGAGTCT GGGGGTAAAT
29851 CCCTGGAGGC CTGTGGCAGC CTCAGAGGTG TGTTCCTCCT GCACTTTCTT
29901 CAAGAAGAAT GTGGCTGCC CTGCTGAGCC TCGTTCGCCC CGTTCGCCCC
29951 GGGCAGTCCC GGCCAAATGTC AGCGCAGCAA GGGGAGGGCC TCTGTAACCA
30001 GGGCTGCTGG CTGCGGGGCT CCCCACTGGA CACGGGAGCG GACATTGGAG
30051 TGTCTTCAT CCGCGTCACT CTTCAGGTC CTTGCCTCC CATTTTCCTT
30101 CCTTCCTTCC TCTTTCTTTC CTTCCTCCT TTCTCCTTCC GCTCCCTTCC
30151 TTTTCTTTCT TCCTTCGCTT TTTCTTTCCT TTCTTCCTTC CTTCTTTTTT
30201 TTCCTTCCCG TTCTCCCTTT CTTCCTCCT CCACCCCTTT CTTCCTACTG
30251 TGTTCCCCAG GCTGGTCTTC AACTCCTGGG CTCAAGCAAT CCTCTTGCCT
30301 CAGCCTCTGG TGTGGCTGGG ACCACAGACA CATGCCACCG CGCCAGGCTA
30351 ATTTGTTAAT TTTTTTTATA GAGACAGGGT CTCACCTTGT TTCCAGGCT
30401 CGTCTCAAAC TCCTGGGCTC AAGTGATACT CCAGCTTCAG CCTGAAGTGC
30451 TGGAATTAAG GTGTGAGCCA CCATGCCTGG CCCTCTCTCA TTCACAAGTG
30501 AACCGTTTAC CCCTGCCCTC CAATCCATGT CGTTTCTGAC CTCAGGCAGC
30551 TTCTCTTTCT ACATAAAGTG AACCTGTCCC AAAGCTGTGC TCACTGGGCC
30601 CCCCTGCCAG GGCTGGAAGA GGCAGCAGTT CACATTTGGC TTGCACTCAC
30651 ATACCAAGGG CATCCATGCT TGAACCTAGA CATGGTTCAT TCACAGGGAT
30701 GGGTGAAGGT AAACAGACTG TGGCGGGCAC TAGGCACTAT CAGTTTCATA
30751 AATCTGCATA CCACCTGTGA CTCAGCAGTT ATGCTTCTCG GAATCTACTC
30801 AAACATGCTT GAGTCAGCCT TCCAGGAAGC AGGCACAAGG ACGTTTGTGA
30851 CAACTTGGTC AGTAATTTTA AGAAGTAGGA AGAAACCTGA GTTTCTCTGA
30901 ATTCCGTAAC ATCTTGACTG TAGGACACAC GAATAATGCC GTGGAATATT
30951 GTGCAGCAGT TAAAAGAAAT GAGGTGAAGG CCGGGTGTGG CAGTCACGC
31001 CTGTTATCTC AGCACTTTGG GAGGCCGAGG CTGGTGGATC ACTTAAGGTC
31051 AGGAGTTTGA GACCAGTCTG GCCAACATGG CGAAACCTG TCTCTACTAA
31101 AAATACAAAA ATTAGCTGAC CTCTGTGACA GAGTGAGACT CTGCCTCAAA
31151 AAAAAAAAAA AAAAAAAAAA AAAAATGAGG TGAAGCTTTA AGTTGTAACA
31201 CTGATTTTGG GGTGCAATAA AGCAAGTTGC AGAATGATAC CCATGTTAAG
31251 ATGCTATTGG AGTGAACACA CGGACCAAAC AATTCTATGT TGGGTACAAA
31301 TTTAAGAAAG AGTTCGGGCG TGGGCACGGT GGCTCACAAC TGTAAATCCCA
31351 GCACTCTGGG AGGGTGAGTC GGATGGATCA GGAGGTGAGG ACTTCTAGAC
31401 TAGCCTGGCC AATATGGTGA AACCCCATCT CTAATAAAAA TACAAAAATT
31451 AGCCAGGCGT GTTGGCGCGT GTCTGGAGTC CCAGCTACTC AGGAGACTAA
31501 GGTAGGAGAA TCACTTGAAC CCGGGAGGCA GAGGTTGCAG TGAGCTGAGA
31551 TCATGCCACT GCACTCTGGC CTGGGCAACA GAGCGAGACT CAAAAAATAA
31601 ATAAATAAAT AAATAAAATA AATAAGAGTT TTCTGCACTT TGGGAGGCCT
31651 GTAGTCCCAG CTACTCTGGA GGCTGAGGCA AGAGGATCAC TTGAGCCTGG
31701 GGGGGTCGAG GCTGCAGTGA GTCCTGATTG TGTCACTGAA ATCCAGCCTG
31751 GGCAACAGAG TGAGAC (SEQ ID NO:5)

```

FEATURES:

```

Start:          914
Exon:           914-1147
Intron:         1148-2008
Exon:           2009-2286
Intron:         2287-2517
Exon:           2518-2563
Intron:         2564-5066
Exon:           5067-5146
Intron:         5147-7047
Exon:           7048-7190
Intron:         7191-15250
Exon:           15251-15426
Intron:         15427-15844
Exon:           15845-15917
Intron:         15918-18499
Exon:           18500-18550
Intron:         18551-20881
Exon:           20882-20942
Intron:         20943-21839
Exon:           21840-21895
Intron:         21896-22194
Exon:           22195-22335

```

Intron: 22336-23816
 Exon: 23817-23944
 Intron: 23945-24261
 Exon: 24262-24361
 Intron: 24362-24840
 Exon: 24841-24974
 Intron: 24975-27018
 Exon: 27019-27091
 Intron: 27092-28484
 Exon: 28485-28576
 Stop: 28574

CHROMOSOME MAP POSITION:

#	SHGCNAME	CHROM#	LOD SCORE	DIST. (cRs)
1	SHGC-56719	17	15.9	14
2	SHGC-58932	17	15.68	14
3	SHGC-97	17	14.5	17
4	SHGC-53147	17	13.3	22
5	SHGC-33067	17	12.52	23

ALLELIC VARIANTS (SNPs):

POSITION	Major	Minor	Context
30344	g	a	aagcaatcctcttgccctcagcctctggtgtggctgggaccacagacacat [g/a]ccaccgcgccaggctaatttgttaatttttttatagagacagggtctca
31170	a	g	ctctgtgacagagtgtgagactctgcctcaaaaaaaaaaaaaaaaaaaaaa [a/g]aaaatgagggtgaagctttaagttgtaacactgattttgggggtgcaataaa
16256	c	t	aaacgctgctgtctgattttaaggtaacgtcactgaaaggggagtttgca [c/t]atggagctgggttgagatctgcatgaacaatcatattctatggtgtctcc
13376	a	t	catgccagctaattttgtatttttagtaaaagacagagcttcactgtgt [a/t]ggctggggcagctctcgaaattcctgacttcagggtgattcagctgcctcggc
12210	t	c	atgttgggattacagggtgtgagccactactcctggctcttcttcaacttt [t/c]gacttttgtatgtctgaaagtctattttgccttcatttttcaaagatagt
12072	g	c	catgcatcaccacgcctggataatttttatttttactttttgtagagat [g/c]gtgtcttctacgttgcccaggctggtctcttactcctgggctcaagga
11922	g	t	taactctttcagcatttttttttttttttttttttttttttttttttttttt [g/t]tactcaggctggaatgcagtggtacaattatagttcactgcagccttga
11903	-	a	ttgcaggcctgttggtgattaactctttcagcatttttttttttttttttt [-/a]tttgacaggattcactctgtcactcaggctggaatgcagtggtacaatta
10009	c	g	ctttgacctgtggattatttagaagtggtatgatgatttcttctgtgaa [c/g]catgggagttatttagtttccaaatagttgggggtttcccaattagtact
4519	c	t	cctgccagagacctcagagctggtatcgtgggtggaggctcctactttgc [c/t]gaggattccccaaagctggtttcttgaagccctcagagccctccacatct
4181	a	g	aacagtcccagagtccgccccccaggatgtcctaaccaccaaccacctgt [a/g]cgagcagttttagggcagggtccttgagcctgtggccatgggatccagg

20952	a	c	tccctcgtctcctcctacccgggtcacaggcagctttggacggtgagtgac [a/c]tgccgcctctctgtttgccacggttgacgccttaacggtgttacgct
20987	t	c	tggacggtgagtgacatgtccgcctctctgtttgccacggttgacgcc [t/c]taacggtgttacgctgacaaggagtgctgcctgcctgaccccggcgcgcc
21620	g	a	ccccacctggatggggcaggagacagagtcggcaggaacctgaaaggac [g/a]cgtgcttctgagcttcttctatagtcaggggtggcccaagcgcggtgt
21795	t	c	ctgacgtccctcggagatcagccacaggagtggtgactgaggtctccc [t/c]ttcccgccctggtgactgacggtctctgtgttgccctccaggacagc
22753	a	t	agcgaacagccttgaccctggctcagaatggcagttccttttttttt [a/t]ttattattatttttttttttttttttattgatcattcttggtgtttcttgc
22945	g	a	gcagaacaaaatgaaaagtctcccatgtctacttctttctacacagacac [g/a]gcaaccatccgatttctcaatcttttccccacctttccccctttctatt
23032	g	a	ccccctttctattccaaaaaccgccattgtcatcatggcggttctca [g/a]tgagctgttggtacacctcccagacggggtggtggccgggcagaggggc
23738	g	a	ggggctgcttggggcccatgagcacctttactcatatgtggggggcagaa [g/a]gctgtcccgtggtcagcagggccatgttggggcctcgggcagctgccgg
23952	t	g	tctgctcatgctcctgcactctgcagccaggcctgagaccaaggtacccc [t/g]ccgtggcctctgagtggggagtggtgctgggggcaggattcctgggcatgg
24123	a	g	gggcagtcaccttgctataaaaccatggtgttctccactgtgtgggggcc [a/g]tgggggtctccccttagcacccctctcccggtcccctgcagcacgctagg
24527	c	-	caggcatctctgagtggtggctggaccgtcctctgtggggcctcagccagtgg [c/-]tgctgcagcaagggtggtggctccccacatatcactccttcctgccccct
24691	c	t	agggaaacagctgctgctgtcaccaacaattgtctctggtcctgccacc [c/t]gaatcccccaactgggcgactcagccgccacgagatggagcactctggcc
25015	g	a	gggcctgcagggtgggtgtgcactgggatgccttaggggttagcagctgcc [g/a]gaaggccttctgtgcctgcctcccatggcgaatgtgacatctctgggct
25191	g	t	ctcgtccctccctgtggaagggggagcggtggccccagccctccgaggg [g/t]tcacgttatggcttctggtcactgccacactgtcctttgtgctggggaca

POSITION	Major	Minor	
30344	g	a	Intron
31170	a	g	Intron

16256	c	t	Intron
13376	a	t	Intron
12210	t	c	Intron
12072	g	c	Intron
11922	g	t	Intron
11903	-	a	Intron
10009	c	g	Intron
4519	c	t	Intron
4181	a	g	Intron
20952	a	c	Intron
20987	t	c	Intron
21620	g	a	Intron
21795	t	c	Intron
22753	a	t	Intron
22945	g	a	Intron
23032	g	a	Intron
23738	g	a	Intron
23952	t	g	Intron

24123	a	g	Intron
24527	c	-	Intron
24691	c	t	Intron
25015	g	a	Intron
25191	g	t	Intron

Gene structure of Splice Form 2:

protein 1 MPSSVTALGQARSSGPGMAPSACCCSPAALQRRLPILAWLPSYSLOWLK
 MPSSVTALGQARSSGPGMAPSACCCSPAALQRRLPILAWLPSYSLOWLK
 MPSSVTALGQARSSGPGMAPSACCCSPAALQRRLPILAWLPSYSLOWLK
 genomic 687 acttgagcgcgattgcgagcagttttcgccaaccacgtccattctca
 tccctcctgacgcgcgtccgcgggccccctaggtcttcgtcgactagta
 gttgggggtgcgctccgcgcgccccctgcgggggcggggccccggggg

protein 50 MDFVAGLSVGLTAIPQALAYA EVAGLPPQ
 MDFVAGLSVGLTAIPQALAYA EVAGLPPQ
 MDFVAGLSVGLTAIPQALAYA EVAGLPPQ
 genomic 834 agtgggctggcagaccgcgtgggggccccGTGAGGC Intron 1
 tattcgtctgtcctcactcacatcggtcca<0-----[921 : 1781]
 gtcccccatcctctcgggcttagtacgcg

protein 79 YGLYSAFMGCFVYFFLGTSRDVLGPTAIMSLLVSFYTFHEPAYAV
 YGLYSAFMGCFVYFFLGTSRDVLGPTAIMSLLVSFYTFHEPAYAV
 YGLYSAFMGCFVYFFLGTSRDVLGPTAIMSLLVSFYTFHEPAYAV
 genomic 1779 CAGtgcttgtagttgtttcgatcggacgcagaatccgtttatcgcggtgg
 -0>agtaccttggttatttgcgatctgcccttctttctactaaccact
 tccctccgcccgtccgcccgtgtgccctgcgcgccccctgcacctg

protein 125 LLAFLSGCIQLAMGVRLRLAHISPHPLGLGGAGTSSMSPLGWPGFLLDFI
 LLAFLSGCIQLAMGVRLRL P G L L F
 LLAFLSGCIQLAMGVRLRLGEALPSCQG-----DALDLSICLFAFQ
 genomic 1920 ccgtctgtaccgagcctggccttcg ggcgcaatttgtc
 ttcttcggtatctgttgtgactccgag actatgtgttcta
 ggccgccccggcgcggtgtgtattcag tccccctcgata

protein 174 SYPVIK FTSAAAVTIGFGQIK
 YP ++ FTSAAAVTIGFGQIK
 VYPRVR G:V[gtc] FTSAAAVTIGFGQIK
 genomic 2040 gtccgcGTGTGTGCG Intron 2 AAGCtatggggaagtgcaa
 tacgtg <2-----[2060 : 2347]-2> tccccctctgtgata
 ctctgt ctttccccctagcg

protein 196 NLLGLQNI PRPFFLQVYHTFLRIAET
 NLLGLQNI PRPFFLQVYHTFLRIAET
 NLLGLQNI PRPFFLQVYHTFLRIAET
 genomic 2394 GTAGGCA Intron 3 CAGaccgccaacacttccgtcatcaagga
 <0-----[2394 : 4901]-0>attgtaatcgctttataacttgtcac

cggaagcccggccgggcccccgtagc

protein 222 VGDAVLGLVCM L L L L L V L K L M R D H
VGDAVLGLVCM L L L L L V L K L M R D H
R:R[agg] VGDAVLGLVCM L L L L L V L K L M R D H
genomic 4980 AGGTACCCC Intron 4 CAGGggggggcgctacccccgcacacgc
<2-----[4982 : 6882]-2> tgacttggtgtttttttattgaa
atccccgggcccggggggggggggcc

protein 246 VPPVHPEMPPGVRLSRGLVWAATT
VPPVHPEMPPGVRLSRGLVWAATT
VPPVHPEMPPGVRLSRGLVWAATT A:A[gct]
genomic 6953 gccgccgaccggccacgcgtggaaGGTGAGGG Intron 5
tcctacatccgtgtgtgggtgcccc <1-----[7026 :10803]
gtccccggccttgccctggcgtcga

protein 270 RNALVVSFAALVAYSFEVTGYQPFILTGETAEGLPVRI P P F S V
RNALVVSFAALVAYSFEVTGYQPFILTGETAEGLPVRI P P F S V
RNALVVSFAALVAYSFEVTGYQPFILTGETAEGLPVRI P P F S V
genomic 10801 TAGCTcagcggttgccggtttggagtcctacaggagggcccgccaccttg
-1> gactttctccttcactatcgaactttcgaccagtcctgtcctct
cccggcccacgtgcccggtacgtccaaggatggctacgcgccag

protein 315 TTANGTISFTEMVQ DMGAGLAVVPLM
TTANGTISFTEMVQ DMGAGLAVVPLM
TTANGTISFTEMVQ DMGAGLAVVPLM
genomic 10938 aagagaattagagcGTGGGCG Intron 6 CAGgagggcgggcca
cccagctctcatta<0-----[10980:11397]-0>atgcgtcttctt
caccggccccgggg cgacggcgggcgg

protein 341 GLLESIAVAKAF SQNNYRIDANQ
GLLESIAVAKAF SQNNYRIDANQ
GLLESIAVAKAF A:A[gca] SQNNYRIDANQ
genomic 11434 gccgaaggaggtGGTAAGAC Intron 7 TAGCATcaatcaggac
gttagtctcact <1-----[11471:14777]-1> caaaagtacaa
ccggctggcacc tgttccctccg

protein 365 ELLAI LTNMLGSLVSSYPVTGSF
ELLAI LTNMLGSLVSSYPVTGSF
ELLAI G:G[ggg] LTNMLGSLVSSYPVTGSF
genomic 14813 gccgaGGTAAGAC Intron 8 CAGGTcaaatgtcgtttcgagat
attct <1-----[14829:15487]-1> tcattgcttcactcggg
gggcc cccggccccccgcacct

protein 389 G TAVNAQSGVCTPAGGLVT
G TAVNAQSGVCTPAGGLVT
G TAVNAQSGVCTPAGGLVT
R:R[cgg] TAVNAQSGVCTPAGGLVT
genomic 15544 gCGGTGAGTG Intron 9 CAGGaggagctgttacgggcga
g <2-----[15549:16445]-2> cctacacgtgcccggttc
a acgctggggccggggcggg

protein 409 VLVLLSLDYLTSLFYIIPKSALA
VLVLLSLDYLTSLFYIIPKSALA
G:G[gga] VLVLLSLDYLTSLFYIIPKSALA
genomic 16501 GGTAAGGC Intron 10 CAGGAgcgccctcgatctttacatgcg
<1-----[16502:16800]-1> tttttctaataccttaatacacctc
gggggtgccgcagcccccgctcgt

26/27

intron	7026	10803
cds	10804	10979
intron	10980	11397
cds	11398	11470
intron	11471	14777
cds	14778	14828
intron	14829	15487
cds	15488	15548
intron	15549	16445
cds	16446	16501
intron	16502	16800
cds	16801	16941
intron	16942	18422
cds	18423	18550
intron	18551	18867
cds	18868	18967
intron	18968	19446
cds	19447	19580
intron	19581	21620
cds	21621	21693
intron	21694	22916
cds	22917	23005

SEQUENCE LISTING

<110> APPLERA CORPORATION

<120> ISOLATED HUMAN TRANSPORTER PROTEINS,
 NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
 AND USES THEREOF

<130> CL000746PCT

<140> TO BE ASSIGNED

<141> 2001-10-26

<150> 09/765,344

<15> 2001-01-22

<160> 7

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 2919

<212> DNA

<213> Human

<400> 1

aacagcacga gggcggaccc agctgtggcg acgccaggag accccaagct gcatcgccga 60
 gtggaagcaa ctagaactcc agggctgtga aagccacagg tgggggctga gcgaggcgtg 120
 gcctcaggag cggaggaccc cccactctc cctcgagcgc cgaggtccac cgtagcgggt 180
 ggagcccgcc ttggtgcgca gttgaaaac ctggagccc cgctggatct cctggctgcc 240
 accgcacccc cccgccagcc tacgcccac cgtagagatg ccttctcgg tgacggcgct 300
 gggtcaggcc aggtcctctg gccccgggat ggcccagagc gcctgtgtgt gctcccctgc 360
 ggccctgcag aggaggctgc ccatctggc gtggctgccc agctactccc tcagtggtgt 420
 gaagatggat ttctgcgcc gcctctcagt tggcctcact gccattcccc aggcgtggc 480
 ctatgtgtaa gtggctggac tccgcccga gtagggctc tactctgcct tcatgggctg 540
 ctctgtgtat ttctctcgg gcacctccc ggaatgtgact ctgggcccc cgcgcattat 600
 gtccctcctg gtctcctct acacctcca tgagcccgcc tacgtgtgc tgctggcctt 660
 cctgtccggc tgcatccagc tggccatggg ggtctcgct ttgggggtcc tgctggactt 720
 catttcciac cccgtcatta aaggcttcac ctctgtgtgt gccgtcacca tcggctttgg 780

acagatcaag aacctgctgg gactacagaa catccccagg ccgttctcc tgcagggtga 840
 ccacaccttc ctacaggattg cagagaccag ggtagggtgac gccgtcctgg ggctgggtctg 900
 catgctgctg ctgctgggtgc tgaagctgat gcgggaccac gtgcctcccg tccaccccg 960
 gatgccccct ggtgtgcggc tcagccgtgg gctgggtctgg gctgccacga cagctcgcaa 1020
 cgccctgggt gtcctctcg cagccctggt tgcgtactcc ttcgagggtga ctggatacca 1080
 gcctttcalt ctaacagggg agacagctga ggggtccct ccagtccgga tcccgccctt 1140
 ctacgtgacc acagccaacg ggacgatctc cttaccgag atggtgcagg acatgggagc 1200
 cgggtggcc gtggtgcccc tgatggcct cctggagagc attgcggtgg ccaaagcctt 1260
 cgcatctcag aataattacc gcacgatgc caaccaggag ctgctggcca tcggctcac 1320
 caacatgttg ggctccctcg tctctcta cccggtcaca ggcagcttg gacggacagc 1380
 cgtgaacgct cagtcggggg tgtgacccc gccggggggc ctggtgacgg gagtgtgtgt 1440
 gctgtgtct ctggactacc tgacctact gttctactac atcccaagt ctgccctggc 1500
 tgccgtatc atcatggccg tggccccgt gttcgacacc aagatctca ggacgtctg 1560
 gcgtgttaag aggtggacc tgctgcccc gtgcgtgacc ttcctgtgt gctctggga 1620
 ggtgcagtac ggcacctcg ccggggccct ggtgtctctg ctcatgtcc tgcactgac 1680
 agccaggcct gagaccaagg tgcagaggg gccgggtctg gtccgcagc cggccagcg 1740
 cctgtcttc cctgcatgg aggtctgctg ggaggagatc ctaagccgg cctggaagt 1800
 gtcccccca cgctgcctgg tctggagt caccatgtc tgcagcatc actacactgt 1860
 ggtgctggga ctggcgagc tctccagga cttcagaag caggcgctg cctggcctt 1920
 tggggcctg caggtcccc ttcctgtgt cctgtgtcc gctgacctg aggggtcca 1980
 gtacttctt accctggaag aagcagagaa gcacctgagg caggagccag ggaccagcc 2040
 ctacaacatc agagaagact ccattctgga ccaaaggtt gccctgtca aggcataatg 2100
 gggccaccg tgggcatcca cagttgcag ggtgtccg aaggttctg tcatgtgat 2160
 tggatgtgg atgccctg atagacatgc tggcctggct gagaacccc tgagcaggta 2220
 acccaggga gagaaggaag ccaggcctgg aggtccacg cagtgggagt ggggtcact 2280
 ggctctctg gggatgactg gaaaatgacc tcgtgtgt tccctggcat gacctctt 2340
 ggaagagtgg ttggagaga gcctctaga atgacagact gtgcaggaa gcaggggcag 2400
 ggggttcag ccgggtgt gcgagcatc ctggggctg cagcacctc ccggtcacc 2460
 agtccacct gcgggggagg gacggggcag gcaggagtct gggaggcgg tccgtctc 2520
 ttgtctcgg catctgtct cccgagaga aaaccaaggt gtgtcaaat acgtcaagt 2580
 tctaittaa aataatttg tgtttctaa atggaaaaa tgatagctt ggtgatttg 2640
 taaaagtc ataatctat tgtaaaaaa acaggaaacc accctcacc ctgtccact 2700
 gggatcatc tccagacccc tcccaaaca tgcataatg cctgtccgtc agtgtlgga 2760
 tgaatglla cagttctaca taaatggat cattttatc atggtgtct ggaaccaca 2820
 ttttcatc agtcattgc agtgaattat ttattgtat aataaatagc attagaatac 2880
 aagaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 2919

<210> 2

<211> 2955

<212> DNA

<213> Human

<400> 2

gctgtggcga cgccaggaga cccaagctg catcgccgag tggaagcaac tagaactcca 60
 gggctgtgaa agccacaggt gggggctgag cgaggcgtgg cctcaggagc ggaggacccc 120
 cccactctcc ctgagcgcc gcagtccacc gtagcgggtg gagcccgctt tggcgcgag 180
 ttgaaaacc tcggagcccc gctggaatc ctggctgcca ccgcacccc ccgccagcct 240
 acgccccacc gtgagatgc ctcttcgtt gacggcgctg ggtcaggcca ggtcctctgg 300
 ccccgggatg gccccgagcg cctgtgtgt ctccctgcg gccctgcaga ggaggctgcc 360
 catctggcg tggctgcca gctactccct gcagtggctg aagatggatt tcgtcgccg 420
 cctctcagtt ggcctcactg ccattccca ggcgtggcc tatgtgaag tggctggact 480
 cccgccccag tatggcctt actctgcct catgggctgc ttcgtgtatt tctcctggg 540
 cactccccg gatgtactc tgggccccac cgccattatg tccctcctg tctcctcta 600
 cacttccat gagcccgctt acgtgtgtc gctggcctc ctgtccgct gcattcagct 660
 ggccatgggg gtctgcgtt tggctacat ctccctcat cctctgggac tgggtggagc 720
 cgggaccagc tcgatgtccc ctctggctg gccagggttc ctgtggact tcattccta 780
 cccgctcatt aaaggctca cctctgtgc tgcgtcacc atcggtttg gacagatcaa 840
 gaacctgtg ggactacaga acatccccag gccgttctc ctgcagggt accacacctt 900
 cctcaggatt gcagagacca gggtaggtga cgccgtcctg gggctggtc gcattgtct 960
 gctgtgtgt ctgaagctga tgcgggacca cgtgcctccc gtccacccc agatgcccc 1020
 tgggtgtcgg ctacccgtg ggtgtgtct ggctgccacg acagctcgca acgcccgtgt 1080
 ggtctcttc gcagccctg ttcgtactc ctgcagggt actggatacc agccttcat 1140
 cctaacaggg gagacagctg aggggctccc tccagtccg atccgcctt tctcagtac 1200
 cacagccaac gggacgatc cctcaccga gatgtgcag gacatgggag ccgggctggc 1260
 cgtgtgtccc ctgatgggc tctggagag caltgcgggt gccaaagcct tcgcatccta 1320
 gaataattc cgcacgatg ccaaccagga gctgtggcc atcggttca ccaacatgtt 1380
 gggctccctc gtctctctt acccggtcac aggcagctt ggacggacag ccgtgaacgc 1440
 tcagtcgggg gtgtgacccc cggcggggg cctggtgacg ggagtgtgtg tctgtgtc 1500
 tctggaclac ctgacctac tgttacta catcccaag tctgccctg ctgccgtcat 1560
 catcatggcc gtggccccg tttcgacac caagatctt aggcagctt ggcgtgttaa 1620
 gaggctggac ctgtgcccc tgtgggtgac ctctctgtg tcttctggg aggtgcagta 1680
 cggcatcctg gccggggccc tgggtctct gctatgtc ctgcactct cagccaggcc 1740
 tgagaccaag gtgtcagag ggccggttct ggtcctgcag ccggccagcg gcctgtcct 1800
 ccctgcatg gaggctctg gggaggagat cctaagccg gccctggaag tgtccccg 1860
 acgtgcctg gtctggagt gcacccatg ctgcagcat gactacact tgggtgtgg 1920
 actggcgag ctctccagg acttcagaa gcaggggct gccctggcct ttgtggcct 1980
 gcaggtcccc gttctcgtg tctgtgtc cgtgacctg aagggttcc agtactctc 2040
 tacctggaa gaagcagaga agcacctgag gcaggagcca gggaccagc cctacaacat 2100
 cagagaagac tccattctg accaaaaggt tgcctgtc aaggcataat gggccaccc 2160
 gtgggcatcc acagttgca ggggttccg gaaggttct gtcactgtg ttgatgtc 2220

gatgccgcct gatagacatg ctggcctggc tgagaaaccc ctgagcaggt aaccagga 2280
 agagaaggaa gccaggcctg gaggtccacg gcagtgggag tggggctcac tggcttctg 2340
 tgggatgact ggaaatgac ctgctgctg ttccctggca tgacctctt tgaagagtg 2400
 gtttgagag agccttctag aatgacagac tgtgagga agcaggggca ggggttcca 2460
 gcccgggctg tgcaggcat cctggggctg gcagcacctt cccggctcac cagtccacc 2520
 tgcgggggag ggacggggca ggcaggagtc tgggaggcgg gtccgctcct ctgtctgcg 2580
 gcatctgtc tctccgagag aaaaccaagg tgtgcaaat gacgtcaagt ctctattaa 2640
 aaataattt gtgtttcta aatggaaaa gtgatagctt tgggtattt gtaaaagta 2700
 taaatgcta ttgtaaaaa tacaggaaac caccctcac cctgtccact tgggtatca 2760
 ttccagacc ctcccaaac atgcatatg acctgtcgt cagtgtgtg atgtatgtt 2820
 acagttctac ataatggga tcatttata catgtgtc tgaaccac attttcatg 2880
 cagtcattg cagtgaatta ttattgtga taataaatg cattagaata caaaaaaaaa 2940
 aaaaaaaaa aaaaa 2955

<210> 3

<211> 606

<212> PRT

<213> Human

<400> 3

Met Pro Ser Ser Val Thr Ala Leu Gly Gln Ala Arg Ser Ser Gly Pro
 1 5 10 15
 Gly Met Ala Pro Ser Ala Cys Cys Cys Ser Pro Ala Ala Leu Gln Arg
 20 25 30
 Arg Leu Pro Ile Leu Ala Trp Leu Pro Ser Tyr Ser Leu Gln Trp Leu
 35 40 45
 Lys Met Asp Phe Val Ala Gly Leu Ser Val Gly Leu Thr Ala Ile Pro
 50 55 60
 Gln Ala Leu Ala Tyr Ala Glu Val Ala Gly Leu Pro Pro Gln Tyr Gly
 65 70 75 80
 Leu Tyr Ser Ala Phe Met Gly Cys Phe Val Tyr Phe Phe Leu Gly Thr
 85 90 95
 Ser Arg Asp Val Thr Leu Gly Pro Thr Ala Ile Met Ser Leu Leu Val
 100 105 110
 Ser Phe Tyr Thr Phe His Glu Pro Ala Tyr Ala Val Leu Leu Ala Phe
 115 120 125
 Leu Ser Gly Cys Ile Gln Leu Ala Met Gly Val Leu Arg Leu Gly Phe
 130 135 140
 Leu Leu Asp Phe Ile Ser Tyr Pro Val Ile Lys Gly Phe Thr Ser Ala
 145 150 155 160

Ala Ala Val Thr Ile Gly Phe Gly Gln Ile Lys Asn Leu Leu Gly Leu
 165 170 175
 Gln Asn Ile Pro Arg Pro Phe Phe Leu Gln Val Tyr His Thr Phe Leu
 180 185 190
 Arg Ile Ala Glu Thr Arg Val Gly Asp Ala Val Leu Gly Leu Val Cys
 195 200 205
 Met Leu Leu Leu Val Leu Lys Leu Met Arg Asp His Val Pro Pro
 210 215 220
 Val His Pro Glu Met Pro Pro Gly Val Arg Leu Ser Arg Gly Leu Val
 225 230 235 240
 Trp Ala Ala Thr Thr Ala Arg Asn Ala Leu Val Val Ser Phe Ala Ala
 245 250 255
 Leu Val Ala Tyr Ser Phe Glu Val Thr Gly Tyr Gln Pro Phe Ile Leu
 260 265 270
 Thr Gly Glu Thr Ala Glu Gly Leu Pro Pro Val Arg Ile Pro Pro Phe
 275 280 285
 Ser Val Thr Thr Ala Asn Gly Thr Ile Ser Phe Thr Glu Met Val Gln
 290 295 300
 Asp Met Gly Ala Gly Leu Ala Val Val Pro Leu Met Gly Leu Leu Glu
 305 310 315 320
 Ser Ile Ala Val Ala Lys Ala Phe Ala Ser Gln Asn Asn Tyr Arg Ile
 325 330 335
 Asp Ala Asn Gln Glu Leu Leu Ala Ile Gly Leu Thr Asn Met Leu Gly
 340 345 350
 Ser Leu Val Ser Ser Tyr Pro Val Thr Gly Ser Phe Gly Arg Thr Ala
 355 360 365
 Val Asn Ala Gln Ser Gly Val Cys Thr Pro Ala Gly Gly Leu Val Thr
 370 375 380
 Gly Val Leu Val Leu Leu Ser Leu Asp Tyr Leu Thr Ser Leu Phe Tyr
 385 390 395 400
 Tyr Ile Pro Lys Ser Ala Leu Ala Ala Val Ile Ile Met Ala Val Ala
 405 410 415
 Pro Leu Phe Asp Thr Lys Ile Phe Arg Thr Leu Trp Arg Val Lys Arg
 420 425 430
 Leu Asp Leu Leu Pro Leu Cys Val Thr Phe Leu Leu Cys Phe Trp Glu
 435 440 445
 Val Gln Tyr Gly Ile Leu Ala Gly Ala Leu Val Ser Leu Leu Met Leu
 450 455 460
 Leu His Ser Ala Ala Arg Pro Glu Thr Lys Val Ser Glu Gly Pro Val
 465 470 475 480

Leu Val Leu Gln Pro Ala Ser Gly Leu Ser Phe Pro Ala Met Glu Ala
 485 490 495
 Leu Arg Glu Glu Ile Leu Ser Arg Ala Leu Glu Val Ser Pro Pro Arg
 500 505 510
 Cys Leu Val Leu Glu Cys Thr His Val Cys Ser Ile Asp Tyr Thr Val
 515 520 525
 Val Leu Gly Leu Gly Glu Leu Leu Gln Asp Phe Gln Lys Gln Gly Val
 530 535 540
 Ala Leu Ala Phe Val Gly Leu Gln Val Pro Val Leu Arg Val Leu Leu
 545 550 555 560
 Ser Ala Asp Leu Lys Gly Phe Gln Tyr Phe Ser Thr Leu Glu Glu Ala
 565 570 575
 Glu Lys His Leu Arg Gln Glu Pro Gly Thr Gln Pro Tyr Asn Ile Arg
 580 585 590
 Glu Asp Ser Ile Leu Asp Gln Lys Val Ala Leu Leu Lys Ala
 595 600 605

<210> 4

<211> 630

<212> PRT

<213> Human

<400> 4

Met Pro Ser Ser Val Thr Ala Leu Gly Gln Ala Arg Ser Ser Gly Pro
 1 5 10 15
 Gly Met Ala Pro Ser Ala Cys Cys Cys Ser Pro Ala Ala Leu Gln Arg
 20 25 30
 Arg Leu Pro Ile Leu Ala Trp Leu Pro Ser Tyr Ser Leu Gln Trp Leu
 35 40 45
 Lys Met Asp Phe Val Ala Gly Leu Ser Val Gly Leu Thr Ala Ile Pro
 50 55 60
 Gln Ala Leu Ala Tyr Ala Glu Val Ala Gly Leu Pro Pro Gln Tyr Gly
 65 70 75 80
 Leu Tyr Ser Ala Phe Met Gly Cys Phe Val Tyr Phe Phe Leu Gly Thr
 85 90 95
 Ser Arg Asp Val Thr Leu Gly Pro Thr Ala Ile Met Ser Leu Leu Val
 100 105 110
 Ser Phe Tyr Thr Phe His Glu Pro Ala Tyr Ala Val Leu Leu Ala Phe
 115 120 125

Leu Ser Gly Cys Ile Gln Leu Ala Met Gly Val Leu Arg Leu Ala His
 130 135 140
 Ile Ser Pro His Pro Leu Gly Leu Gly Gly Ala Gly Thr Ser Ser Met
 145 150 155 160
 Ser Pro Leu Gly Trp Pro Gly Phe Leu Leu Asp Phe Ile Ser Tyr Pro
 165 170 175
 Val Ile Lys Gly Phe Thr Ser Ala Ala Ala Val Thr Ile Gly Phe Gly
 180 185 190
 Gln Ile Lys Asn Leu Leu Gly Leu Gln Asn Ile Pro Arg Pro Phe Phe
 195 200 205
 Leu Gln Val Tyr His Thr Phe Leu Arg Ile Ala Glu Thr Arg Val Gly
 210 215 220
 Asp Ala Val Leu Gly Leu Val Cys Met Leu Leu Leu Val Leu Lys
 225 230 235 240
 Leu Met Arg Asp His Val Pro Pro Val His Pro Glu Met Pro Pro Gly
 245 250 255
 Val Arg Leu Ser Arg Gly Leu Val Trp Ala Ala Thr Thr Ala Arg Asn
 260 265 270
 Ala Leu Val Val Ser Phe Ala Ala Leu Val Ala Tyr Ser Phe Glu Val
 275 280 285
 Thr Gly Tyr Gln Pro Phe Ile Leu Thr Gly Glu Thr Ala Glu Gly Leu
 290 295 300
 Pro Pro Val Arg Ile Pro Pro Phe Ser Val Thr Thr Ala Asn Gly Thr
 305 310 315 320
 Ile Ser Phe Thr Glu Met Val Gln Asp Met Gly Ala Gly Leu Ala Val
 325 330 335
 Val Pro Leu Met Gly Leu Leu Glu Ser Ile Ala Val Ala Lys Ala Phe
 340 345 350
 Ala Ser Gln Asn Asn Tyr Arg Ile Asp Ala Asn Gln Glu Leu Leu Ala
 355 360 365
 Ile Gly Leu Thr Asn Met Leu Gly Ser Leu Val Ser Ser Tyr Pro Val
 370 375 380
 Thr Gly Ser Phe Gly Arg Thr Ala Val Asn Ala Gln Ser Gly Val Cys
 385 390 395 400
 Thr Pro Ala Gly Gly Leu Val Thr Gly Val Leu Val Leu Leu Ser Leu
 405 410 415
 Asp Tyr Leu Thr Ser Leu Phe Tyr Tyr Ile Pro Lys Ser Ala Leu Ala
 420 425 430
 Ala Val Ile Ile Met Ala Val Ala Pro Leu Phe Asp Thr Lys Ile Phe
 435 440 445

Arg Thr Leu Trp Arg Val Lys Arg Leu Asp Leu Leu Pro Leu Trp Val
 450 455 460
 Thr Phe Leu Leu Cys Phe Trp Glu Val Gln Tyr Gly Ile Leu Ala Gly
 465 470 475 480
 Ala Leu Val Ser Leu Leu Met Leu Leu His Ser Ala Ala Arg Pro Glu
 485 490 495
 Thr Lys Val Ser Glu Gly Pro Val Leu Val Leu Gln Pro Ala Ser Gly
 500 505 510
 Leu Ser Phe Pro Ala Met Glu Ala Leu Arg Glu Glu Ile Leu Ser Arg
 515 520 525
 Ala Leu Glu Val Ser Pro Pro Arg Cys Leu Val Leu Glu Cys Thr His
 530 535 540
 Val Cys Ser Ile Asp Tyr Thr Val Val Leu Gly Leu Gly Glu Leu Leu
 545 550 555 560
 Gln Asp Phe Gln Lys Gln Gly Val Ala Leu Ala Phe Val Gly Leu Gln
 565 570 575
 Val Pro Val Leu Arg Val Leu Leu Ser Ala Asp Leu Lys Gly Phe Gln
 580 585 590
 Tyr Phe Ser Thr Leu Glu Glu Ala Glu Lys His Leu Arg Gln Glu Pro
 595 600 605
 Gly Thr Gln Pro Tyr Asn Ile Arg Glu Asp Ser Ile Leu Asp Gln Lys
 610 615 620
 Val Ala Leu Leu Lys Ala
 625 630

<210> 5

<211> 31766

<212> DNA

<213> Human

<220>

<221> misc_feature

<222> (1)...(31766)

<223> n = A,T,C or G

<400> 5

ggtccccgcg gccctcggcc ttgctcgggg ccaagggaacc gcggacggtc aggtggcgca 60
 gggctcctc cggagacccc aggatccgga gccagcggcc ttgtgggcag gggccggggg 120
 caggggagtg gattttgccc ggagcggagc aggccggggg cagtgggggg ctgggggtga 180

gggtggctgg ctctgcgcgc gggcgccggg gccctggaag atgctgcgca cctgaattaa 240
ccggggcgct ctgatgtcct cccagaagca actagaactc cagggctgtg aaagccacag 300
gtgggggctg agcgaggctt ggcctcagga gcggaggacc ccccccccc cccctcgag 360
cgccgcagtc caccgtagcg ggtggagccc gccttggtgc gcagttgaa aacctcgaa 420
gccccgctgg atctcctggc tgccaccgc acccccgcc agctacggtg cgcccgagg 480
ccagcttct ctctgcgtg ctccccgta aattccctgg ggagacggaa aaaaaggaa 540
aggaagtcgg ttctccaggg gccagaagtg ttgagcctaa ttagtctca gacttctca 600
tgaggaatcg cttatcagtt tctatctgg gagagttgag gatggaggga cagaaggcac 660
ccaggatttg cacggggggg gattcaggga gagagggtga tgaggacgg ggtgggcctt 720
ccagcttgg cccagtcctc atctgcaca cattgttgc ttctcttag agccgttcgc 780
ccccctggg aggggagacc catagtacc tctctgaca ccgcccacc ctgaccagt 840
ttgccgggtt ctcaaaggc cagctctga ctgctggtc gtgtacctg cccccccag 900
ccccaccga gagatgcctt ctctggtag ggcgtgggt caggccaggt cctctggccc 960
cggtatggcc ccgagcgct gctgctgct cctgcggcc ctgcagagga ggctgccat 1020
cctggcgtgg ctgccagct actccctga gtggctgaag atggattcg tcgccgcct 1080
ctcagttggc ctactgcca ttcccaggc gctggcctat gctgaagtgg ctggactccc 1140
gccccagggt aggcgtctga cctgctgcc agccatact cagaaacagt gcagaatca 1200
cagatcaat cccagacacc atcagcatt ccaggttcc agccccggg cccaaggaa 1260
ccttggtt acagtgtg acgcagatt tctctggcc gaccaggct cctatgcctg 1320
ttgttacac acagacactg agctggtat ggaggggcca gcgagatgac tcatggaggc 1380
ctcaggagt caagaccagc ccgacaaaaa tggtaaaacc ccgtctccac taaaaataca 1440
aaaattaggc tgggtcggt ggctcaagcc tgtaatcca gcacttggg aggccgaggc 1500
aggcggatcg caaggtcagg agatggagac catcctggct aacagggtga aacctgtct 1560
ctactaaaaa taaaaaaat tagccagggt tggggcggg tgctgtagt cccagctacc 1620
tgaggaggct aggcaggaga atggcgtgaa cccgggaggc ggagcttga gtgagccgag 1680
actgtgccac tgcctccag cctgggcgac agagcgagac tcatctcaa aaaaaaaaaa 1740
aaaagaatgc ttctcagac ttggacacag cacacgggcc tgtaccgacc cctctgcctg 1800
gctgtctga cctgaggcc ccagttgagt gctgctaaa aagtggcctc ctgatcactg 1860
caggtccacc cacagggcag ggcgtgcac cttaacctg ggcctggaca cagctgacac 1920
ccacacatcc cgagcttga cagcacact agggagctgg tggatggcc tcggcctct 1980
gagtgtcac caccctct cccacagta tggcctctac tctgcctca tgggtgctt 2040
cgtgtattc ttctgggca cctccggga tgtgactct ggccccacc ccattatgc 2100
cctcctggtc tcttctaca cctccatga gccgcctac gctgtgtgc tggccttct 2160
gtccggtgc atccagctg ccatgggggt cctgcgttg ggtgaggctc taccttctg 2220
ccaaggggat gccctgacc tcagcattg ctgtttgca ttcaagtct atccccgtg 2280
gcgtgtgtg gcgtgtgtg ggggtgtgg tatgtatg tgtgtgtga ggtgggtgg 2340
tggggaggg ggtggggcac ttgctcct agtctactat ttactgatt agaggccagg 2400
acattggaga aagtgcctg tggctcagac cccatatgcc ccccccccc nnnnnnnnn 2460
nnnnnnnnn nnnnnnnnn nnnnnnnnn nntcattc ctacccgct attaaagct 2520
cacctctgt gctgccgta ccatcggtt tggacagatc aaggtaggca cggcggccac 2580

ccagggcact gctcttggc cactgctcgt tggcacaggg atggcgggag caggactgag 2640
 gccagtcctg atccctgtgg ccagtgagcg tctgtctgt tcagattgtc tccatgggt 2700
 caagaagcac gcggtgctct catgggtccc ctgttaataa aatgaccctc ctgggagggg 2760
 tgtcacgtga tggttggatt tcacagcggg taactgggg gccggttaatt ccatccccct 2820
 gctcttggc gagttccgt gccagtgtgc ttggctggct ctgtgacgtg gctctgtct 2880
 cctgcactg ggcacaccca gcaggcccca ccagtcatga gcatgtcgtc agaatttcta 2940
 taggcaaatt atttcccat gccaattag ttagatgggt ttgtttgtg cttttttc 3000
 tggttctgt ttaatatatt taaatgcc agcctttat atttctttt aaaacattta 3060
 taataagtac taaaatcagt catcagctga ggggttaatt tattctgtt ttgctgggt 3120
 glaagttct cgagggtagg aactgtcct ctcccagcca tggccgcca gcallgggt 3180
 ggtgcagtag ggggggtcct agtggggtgt gtgtggagt aagtggccc agttccacag 3240
 atgggacat gcggccctca tggcagacta ggtcacatg ctgcctcctg accctgtgc 3300
 actgcagtc atcttccca agccaggccc tgtccaggc tggcctgaga cagtcttccc 3360
 tgaaggagt accatgagaa gaccaaggac aggagagtgt gtgtgagagt gtgtatgagt 3420
 gtgtgtgt gtgaatagt gagtgtgta gagtgtgagt gggttgagg gagtgtgt 3480
 gtgcgtgt gagtgtgt gagtgtgggt gtgagtgagt gcgtgtgt gagagagt 3540
 gtgagtgat atgagtga gtgtgagagt ggggtgggt gttaggagt gtgtgtgt 3600
 gcgcgcgcgt gtgtgtgt atgagtgt gagtgtgt gagtgtga gagggtgt 3660
 ggggtgtgt gtgcgtgt gtgtgagagt gtgagtggt gagtgtga ataaagtaga 3720
 cacttttgc actcttcta cgtgcaggc actgggcagg acacttcca tgttaattc 3780
 tcaacacacc cgggaggta gattatcat tattccatt tgagagagga gggaccaact 3840
 taggtgtgg tgagcgtat ttgtacgta tctacagca tcttcaggg ttagctggc 3900
 aaatgctgt tcaggcatg gttgttct taatctgaa acatcattt tgggtcaag 3960
 aatgtttt ttaggatcc cagtgtgagt ggagagcga gagggtgag tgaaggcat 4020
 ccttgtca tgcctcata ctggcaacc ctgccccgc ccagggactc tcagccatc 4080
 tggggggagg ggcgtctcc tgacaggccc aggacagaag accctacccc aacagtcca 4140
 gagtccccc ccaggtatg cctaaccaca cccacctgt acgcagcagt ttagggcag 4200
 ggtcctgag cctgtggcca tggatccag ggcctcaact tccccctg actcctgtt 4260
 ggtcaggc gatgtaaaga actaaagtgc aagctgtcc tgcactgtg atccagcta 4320
 cgtgggagg cgagggtgga ggtactgt agccaggag ttgaatcca gcctgggcca 4380
 cagagcaagc aagaccccca cccccaacc ccccgccgt ctctaaagaa aaaaaacta 4440
 aactaaaaca caggatagag tgccttccc tggcaggac ctgagagctg gtagctgtg 4500
 gggaggctcc tactttgcc aggatcccc aagctgttt ctgaagccc ctgagagccc 4560
 tccacatct cacactacga agagatttt cctccgcag cagcggagct ggggggtggc 4620
 gggcaccta gtgaggaga catttcaag ccaatggcag caagggtc cactgcagg 4680
 ggcctggct ctgagtggc tggagccaat ggggtggg ccaggcagcc catggccct 4740
 ggccatcaag ctgtagaacc tgtctcctg ctgtgggtc acctacattg tttttatga 4800
 gaaaagta ttagagaaaa acatatcact gaccagtaa ttgtgaagta ctgtctcca 4860
 cgagggtagc ttgatctcc tgcctaggg ggcgttggga gtggcaggt ggacctga 4920
 gccclaa gaaggggg gaalaagggt tggagcagg gccgggggac actgtctcag 4980

ccctagggga ggtggcgagg gagctggga cagatggcct tggttggga gcatagcctc 5040
tgatcagcat ctctgtgtt ggacagaacc tgctgggact acagaacatc cccaggccgt 5100
tcttctgca ggtgtaccac accttctca ggattgcaga gaccaggtag cccgggctt 5160
gttctccct cctataagga agtctctt tccacacctc ctctccggc cccacactca 5220
gttcccccac ccttggtgac tgctcaaaca ggggtcccca gagcagcccc atcagcagca 5280
cctgcaagct ggcaagaaat gagacttca gtccttgca gacctgcct atcgagacca 5340
gagggttaac aggaggccgg tggctgcgc gcaggggaga gttcaagaag ctctgtctg 5400
gggagctgga gcagggcaat gttctctct ccacaaagct ttctggggg acaggaatgg 5460
ggagagtctg tgtaaagaaa aagaaggcag tgtgtctccc ccaggctgtg attgttaag 5520
gaggagaaca caggcggtgg ggagctaacc cagacagaac gcagggtgtg caggacggc 5580
agggtgagct gtgatgagag atgaggagac ccaagcactt cctcagggtg gtttaagag 5640
cctctgatgg cgggacactg tggctcacac ctgtaatccc agcactttg gaagcccagg 5700
tgggcagatc atcttaggtc aggagtcaa gaccagctg gtcaaatgg tgaactccg 5760
tcttactaa aaatacaaaa attagtggg tgtgatggtg cgtgcctgta gtccagctg 5820
ttcaggaggc tgaggcagga ggattgctt aacctgtgag gcggagggtg cagttagctg 5880
agatcacgcc actgcactcc agccgggtg atagagttag actccatctc cgcctgcc 5940
tgccccccg aaaagaaaag cctctgatga ggggtacctc cctgccagac catccagcg 6000
gaaggcagga tgctctcta cctctctggc tggagggggc tggaggagga aatgatctag 6060
gggagctata gagatggctg cccagtgtg tggcctggag ggagtggggc cgatcctgga 6120
ccatctccc ctcccctg atctgtgccc gaagctccc agcagctgaa ccagctgga 6180
agccagtggg ttcttgctc ttcagaggca ccagtgggca ggggtcggcc agatggggga 6240
gcagtatta cggagcctga ccagtcagag agcgtcacgg gctgtcatgt aaaaggggcc 6300
ttactgaccg ttatgtcta agcctttag aatgtctaga gcagggtgtc cagtctttg 6360
tcttccctgg gccacacagg aagaattgtc ttaaaatata claatgatag ctgatgagct 6420
acaaaaaat cattgaaaga actcataatg tttaagaaa gtttacaat ttgtgtggg 6480
ccacattcaa agccatccca ggctgcacat ggcccacagg ctgtgggtt ggcagacttg 6540
ttctagagcc acctgcagag agcatcagct gctcaggctg gtcccatc cctgtgtca 6600
ctcaccagat cctgttcat agacttagt cccagagagc tacgggagt aaaatgtgag 6660
cacgtcctc tgcatcatc atgaccggg cctctccact gctgtcctg cccaagcctg 6720
tgtacttact gaatttgaa ttgagtttg aagtagaaca gggatgcata ttgcttagga 6780
gaaaacatct cagcctagtc tgttttgg cttgcagagt agcttctaa agttcctgag 6840
cttagttaa gtgtctgta aatgggaaa atattgaaa tgctttatt cctgtgata 6900
ctcactgtg tctgggacca gcagcctgg cattgggct cctgggagc ttctagtgc 6960
ctctcagct aagcagcagc aggacgctg agtcagcagt aggcctgggc cgagctgggt 7020
gggtaccagt cctctgcctg tccacagggt aggtgacgcc gtctggggc tggctgcat 7080
gctgtctgt ctggtgtga agctgatgc ggaccagct cctccgctc accccagat 7140
gccccctgt gtgcggctca gccgtgggt ggtctgggt gccacgacag gtgaggggcc 7200
tctggctgac atcgtatga acctggctg caggtgggg tcaattggg agtcctagt 7260
ccacctagg gattctacg tcattggtct ggggtcact tgagcattgg gacatttaa 7320
acaccacacc aaactctgg acatgtactt ttatttaat taattaatta attattttt 7380

gagacagagt ttgctctgt cgcccaggct ggagtacagt ggtgtgatat tggctcactg 7440
 caacctccgc ctcccagggt catgtgattc tctgcctca gcctcctgag tagctggat 7500
 tacaggcgca tgcaccacg ctgactgat tttgtaltt atagtagaga cggggttca 7560
 ccatgtgac catactggtc taaaactct gacctgagt gatctgcctg cctcgccctc 7620
 ccgaagtgtc gggattacag gcatgagcca ccgccccgac ccaaactcta gaacatttaa 7680
 aactctgagc caaactcggg gggttctgat gcagggtcag agctgagaac tgatcatgg 7740
 ccagcttggc ccgttctcaa cctgaagcag tttgcccc ctggtgacat gtggcaatgt 7800
 tgggaaacat tttgggtgga taccaccggg cgaggcaggt gctggcaact agcgaggagg 7860
 ggccaggaat gccactaagc cctgacggc gcacagcctt ccacagcaaa cagtatccg 7920
 gccc aaaatg tcggcgggct cgagggtggag aaacacagtc tactggcagg tctgggggag 7980
 acagaacaaa ttccgggggag accaacaggc tctgtttgaa ttgtgcagct ttatgggggt 8040
 acaacttcag cttcaggcgc ggagggtgca ggtgagcca caaccaacca ggggtccttt 8100
 gagaaggggg tggtaggggc tgcgtcgtgg aggagcaggg ctgtcagcag gtgaaggacc 8160
 accctggggg aggtggaagc ctcttctcag tgcattccgg atgggcatgg gtcccctggc 8220
 tgagcaccgg gcagggatgc ccagagaaga aggcacaccc aggaagtca ggaagcatag 8280
 gtgactcaa acctctcag agccgagggg acagattaag ctcaagaata cccgagcttg 8340
 gtctgtctgt ttgggggctg caaaatccaa cagccacaga agagagaggg tggctggatt 8400
 tagctcttga gtcacttctg cctgccagg ctggtctgtt ctctgtggtc tctccctgga 8460
 tctgttcag cagacttgag gccgtggtga acactgagag tctggtgggg tgcagaggat 8520
 gctgggttta gagtggagat gtccagctc cagtccagc ctggcctggg tttagacttg 8580
 caaagtagct tcaaagttc ctgagcttta gtacagttgt ccgtgaaata gtgaaaatgc 8640
 tttgtatct ttgtatctt ttgtttttg tttttttt gagacgggtgt ctactcctg 8700
 ttgccaggc tggagtgcac tggccgtgat ccttggtcga ctgcaacctg cctcctggtt 8760
 ttcaagcaat tattcccc ttaacctcc cgagtagcc tgggatttac aggcctgca 8820
 cccaccacac ccgggcta ttttgtatt tttagttca agatgggggt ttcaccattg 8880
 ttggcccatg ctgggtctcg aaactccnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 8940
 nnnnnnnnnn nnnnnnnncc ctacccctga aaggcaggt ctccgcagcc agaaaggtct 9000
 ttaagatgt ttagacatcg gtaatgtcca gaaacttcag aaacatttta cagtgtaggg 9060
 ctgctgactc ctttctccc agtgtgtgta gcattgtctt aggaactac ttaaaaagct 9120
 ttgatctt ataacagct tgagaatgag ttattattt ttccatttt gtcagtggg 9180
 aaagttagcc tatgaaagat gaaaaataaa acagaacttc tcaaggccac acgactggga 9240
 agaggcagaa caaggactc acccgagtcg tctgtcttt ttctgaatg gtcgtctag 9300
 aagtgtgca atgtatgat cctctcaaag aaccagctt tggttcatt gatctccctt 9360
 ttgcttggg ttctagttc attgattct gctgggctt ttattttt ctttctcta 9420
 ttattttca gtttaactta tctttttt tttttttt aaaaaaaat tttttttt 9480
 gagacggagt ttgctctgt tccccagggt tggagtgcaa tgggtgac tcagctcact 9540
 gcaacctccg cctccgggt tcaagtgtt ctctgcctc agcctccaa gtagctggga 9600
 ttacaggcgt atgccaccat gcctggctaa tttttgtat ttttaataga gacgggggtt 9660
 caccatgtg gccaggctgg tctgaactc ctgaactcag gtgatccacc cgcctcagcc 9720
 tcccaaagt ctgggattac aggcgtgagc cactgtgccc ggctttttt ttacagattct 9780

taaggaggaa gcctgggtct ttgattgag acccttctc ttttlaata tgggtgtta 9840
 aatgctgtaa atttactcc tagcattact ttgctgcat cccacagtt ttatatatat 9900
 gctgtcttt aatttttatt cgggtctaaa tactttgtga ttccatttt gactctcct 9960
 ttgacctgtg gattatttag aagtggatta tatgattct tctgtgaagc atgggagta 10020
 tttagttcc aaatagtgg gggtttcca attagtact tgccattgat ttctaattta 10080
 gttccattgt gataagaaaa cgtacttagt gatttaaalc tgtttcctt tattgagagt 10140
 tgtattatgg ccttgaatat ggtctatctt agtgaatatt ctgtgtgtgc ttgagaacat 10200
 gttttctgcc attattgggt ggaatgtct gtaaatgca gttaggctg gttgggtatc 10260
 tcaggctact atattctgc tgattttct tctactgtt ctgttgatta ttaaaagaag 10320
 aatgtgaaa taactcagc tataaccatg acttgccta tttccttc ttactacca 10380
 gttattgtt catatattt gaagctctgt ttttagtat aaaaacatat agaattataa 10440
 tttctctg gaatgatact tttttctg ggcaatatt tcttccctga aatctactt 10500
 gatactaaga gagctactcc agcttcatt agtgtagca tatctttcc catctttta 10560
 cttttttt ttgagacag aatctctctc tgtgtccag gctggagtgc aatggcgtga 10620
 tctggctca ctgcaacct cacttccag gtcaagcaa ttgtgtgcc tcagcctccc 10680
 aagtagctgg aattacaggc ttgcaccacc aagcccagct aattttgtg ttttagtag 10740
 caatgaggtt tcaccaagt ggccagcctg gtctcaaat ctggcctca agtgatccac 10800
 ctgcctggc ctccaaagt gctgggacac cacaccagc ctctatctt ttactttta 10860
 acgtatttaa ttgatttg ttctgtgt taatggtag tttatgggc agcatataga 10920
 tgggattgt tttaatctc ataactctg cttttaatt ggaatctat gccattata 10980
 ttaagtga ttattggt ctgtgactt agactacca tgtgtgtt tgttctgt 11040
 ttgtctatt tgttctgt tccctttcc ccttctct accctactg gattatttt 11100
 tatgattcca tttaacatc ttattggct gattagctg aacggctata tttagtttt 11160
 gttgttgt ttctaagtga ctgtttaag gtttggtta aacatctta tctgtclaca 11220
 gtataatta aactcaact tcatgtgata taagaacct ataacaggcc aggcgcagtg 11280
 gctcacacct ataactctag cacttggga ggctgaggga ggcagattgt ctgagctcag 11340
 gagtctaaga ccagcctggg caacatggcg aaaccccatc tctacaaaa atacaaaaa 11400
 ttaggcaggc atgggtgtc atgcctgtaa tccagctac tgggaggct gaggtggag 11460
 gattgctga acctggaagg cagagattgc agtgagcaa gatccgtcac tgaagctga 11520
 aactcctgg ctcaagta tctctgcct cagcctctg agtagctgg tctacaggca 11580
 tgcaccacca catctggcta atttttaa tttttata gaaacgggt atgtgtca 11640
 ggctgatcaa aggttctgg cctcaagtga tgcctctgcc tcagccaacc aaagtgtga 11700
 gattgcagat gtgagccacc attcctgca ggaacagct tagattatc cagtagtca 11760
 ctgttctg tctctta tctttgtac aaatccagat ttccatctg tataatttc 11820
 cttctacct aaggatgta ttttcttc tgtgcaggc ctgttggtga ttaactctt 11880
 cagcatttt tttttttt tattgacag gatcactct gtcactcagg ctggaatga 11940
 gtgttacaat tatagttcac tgcagcctg aactcctaga cccaagcaat cctcagcctc 12000
 ctgagtggt gggactatag gcatgcatca ccacgcctgg ataattttt attttactt 12060
 ttgtagaga tgggtctc ctacgtgcc caggctgtc tctactct gggctcaagg 12120
 gatctccca cgagggatcc tccacttg gcttccaaa tgttgggatt acaggtgtga 12180

gccactactc ctggtcttct ttaactttt gacttttga tgtctgaaag tctattttgc 12240
 ctctatttt caaagatagt ttgtctggtt atagaattct agacttttt ttcttttca 12300
 gtactttta aaaaaaagt alttattt ttittatt ttgagacagt tctcgctgtg 12360
 tcaccaggc tggagtgcag tggcacgac ttggctcact gcaatctctg cctcctgggt 12420
 tcaagtgtt ctctgtctc agcctcccag gtagatggaa ttacaggcgc atgctaccac 12480
 gcccggttaa tttgtact tttagcagag atgggggttt gccatgttgg ccaggctgct 12540
 ttccaactc ctgacctcag ggtgatccgc cctcctggc ctccaaagt gctaggatta 12600
 caggcgtag ccacttgcc tggcctcagg acttttaaag atgctattct gctgtcttct 12660
 ggcccactc gttcaagaa gtgtgtggc attcttgggt cctctgtatg taacctttt 12720
 atccaccgt tttaaatat tcttatcac tggtttagg cagttcata atgtggctg 12780
 atgtagtgt tttgtttg ttgttttt ttctctgtg ttatgcttg tgatcattg 12840
 aactcttag acatgtgggt tgcagttc atcaaattg gaaaaaaac acctatctt 12900
 tcgtgaaac ttgtactgt cccctatgc gattccacag actttcttg cagaatcccc 12960
 tgtgtgatt tcaggcacct cgaagggtt ctacagctc actgggatt ttcagcctc 13020
 tccttgtgt gttcatttt ggatagttc tgcgtctaca tttagtctt ttcttctgt 13080
 aatgtctcag ctgttgaat tgcaccaagt gctgttccg tctcagacac tgtggtttc 13140
 attctagaa atccagttg aatctttt ttcttttt ttitttta agatggagt 13200
 tcgtctgt gtccaggctg gagtgcattg gcgcggtct ggctcactg aacctctgc 13260
 tcccggttc aagtgttt cctgcctcag cctcccagg agctgggatt ataggcatc 13320
 accaccatgc ccagctaatt ttgtattt tagtaaagac agagcttcac tgttaggct 13380
 gggccagtct cgaattcctg acttcagggt attcagctgc ctgcgcctcc caaagtgtg 13440
 ggattacagg cgtgagctac cgtgccagc ctcaattga atcttttaa atgtatttc 13500
 tacatctta cttaacagtt gcaatcttc ctctgttt tttaagta tatggaatt 13560
 ggttacagt attctttta tgcclatta actaaattac atttggttac agttactct 13620
 ttaatgtct gtaattctg tcatcttaa ctatgtact ttgggtctt ttittattt 13680
 ttattttta ttttgaga cagggtctca ctgtgtacc caggctagag tacagtggca 13740
 tgatcatggt tcatatagc ctcaacctcc caggctcaag ttatcatct acctagccc 13800
 ccaagtgc tgggactata ggcatagcc accatgtca actaatttt tatttttgt 13860
 ggagacagt tctactaat ttgtcaggc tggcctcaa cttctgggt caagcagtc 13920
 tctgtcttg gcttcccaa gtgtgggat tacaggcatg ggccactgca cccagccata 13980
 ttctctgt tcttgcag cctagtaatt ttgtttgga tggatgctag actttgtgag 14040
 ttgaggta ttgcgtctg gatattttg tattctata atcttctga gctttgtct 14100
 gggatacagt taagtactt aggacagtc gatccattt ggtctactt ttaagcttg 14160
 tgagctggac cagagcagt ttactctag ggcaagccta atttgccca ccacttgatt 14220
 ttgaaalaa agtttatg gaaaatcact cccattcatt tacatgttca atgtcagggt 14280
 ttatgcaata ataacagact tgagtagtg cgccagagac ttatggccc acaaagcaa 14340
 aagtttttag tatttagccc ttacagaaa acatttgcca acccaggcc ttgggcta 14400
 ctccccact gctgaggtaa aacctctgg agtccctgat gctccatga ttatgagggt 14460
 ttctactct ggtgtgaac ctgggattg tccctgtaa cctctttgt gactctccc 14520
 cactgtgggt ttgggtcatt ttgcccagc ttgcctcga ccagttctct gctgcacac 14580

cgagggcacc tctgcagatc tctgcagctc tgtctccggc tgcctctct gctcctgtga 14640
ccttgccggc agctccagct gcaggccctc tcagacttcc tccagcalt gccccctgaa 14700
ctcagggaaa cccctgggct ctgctagggg gtcccccca tacacatggg gaccatcggg 14760
ggaaaccgta ggcagccggc ggggcagtga tagactcacc tccctgttt cctgtggcc 14820
gtcatggctg ctgtcccgcg tcttagtggc catcatgtcc actgtccgc atcttgagg 14880
ccattgttc tgtgtttg gaaaggggtt aaatctggt tcttctct tttgtcttg 14940
agggtgatgt gccctgggcc acctgattc agagagtct tgcgcgggtg cacgatgtcc 15000
agtcacctgg agcttggcag cagggtggat gtgcacctg ccgcagcctg cagctctgcc 15060
gtccaccctg ctatgtctg cacatagcat ttatgtctg gttgtttc ccagagtga 15120
gagtggagac tgtggcccg tggggcgct tctgcctct gggatgtctg cccagtaac 15180
ctggaggcag ccgcgctacc ccacactgc cgggagcagg gtctggacg cacctctct 15240
tctctctag ctgcgaacgc cctgtggct tcttcgcag cctgtgtg gtactcttc 15300
gagggtactg gataccagcc ttcatccta acaggggaga cagctgagg gtcccccca 15360
gtccgatcc cgcctctc agtgaccaca gccaacggga cgatctctt caccgagatg 15420
gtgcagggtg gcggagccgg gaggcaggat gccgtggctg aggtgcagt ggcctctgc 15480
ctggctcta cctgatga tctgtgggt gccaggggt ctgaggtcag ttaggacagc 15540
tgagtctca ggaacggaca tctcagttt taaagaatcc caggttgat gcaaactcag 15600
cgagctcagg gatgtcacgt ttgtttcag gggcgctt cctgtttg actccagctg 15660
aggatgaatt taccgttc ctcccagc ctggcgctc ttacagaaag gaggcggatc 15720
ctgcagctga caagcactg ctctgttac ctgtgggcg ggggtggctc ttgtcttt 15780
catgggtcac tctgtgttc tacccttag gaaggtcact caccatccct ctctcttc 15840
tcaggacatg ggagccgggc tggcgtgtt gccctgatg ggcctctgg agagcattgc 15900
gggtggccaaa gccttcgta agacgcctgt caccacacc ccaggctcc cagtgcgctg 15960
gctgggctag gcctgcctg ttctagctt gcctttatcc gtactagt tttagaaatt 16020
gaattcatat ccaagtaata catgctcatg atagatacat atgtattgt tatatatg 16080
aaaactgat ctataatgag gcatgccctc ccacccatg gtgtgtgtt gagtgtta 16140
acagcctctg ctgtgttg aaataaaagg tttgtctg tggccttc cgatgtccac 16200
gggtgaaacg ctgtgtctg atttaaggt aacgtcactg aaaggggagt ttgcacatg 16260
agctgggtg agatctcat gaacaatcat attctatgt gtctccacca ttagatata 16320
gtgggtgcaa ataactcat cagtagtag caaatgcaa ataaatagg aagtgatgag 16380
tttaagtat tatcttgg ccaggcatg tggccaggc ctgtaatcc aacatttg 16440
gaggctgaga cgggaggatc gcttgagctc aggagtta aaccaccta ggcaacgta 16500
cgagacctg ttctatta aaataaaaat tagctggga tggagcacac ctgtgtacc 16560
agcttctcag aaggctgagg caggagatc actgagccc gggaggtcga ggtggcattg 16620
agctgtgata gtgccactg actccagct gggcagcaga acaagatct gtattatctg 16680
tttaattgaa agtttaatt taattgtaa taatggctat gtttagta atggctcaca 16740
aactctaaa cattagcaa catgcttat tcagctggct cagccatcg gccagccct 16800
cctgcctccc agggaggcag ggaaccctc gtctctca gtgtgttt tagtcagca 16860
tctciagatt cctcctgc agctccgtc ttactcgt ccactctgc cagcttga 16920
ttgtctgc ggaagacacc ttgtctgt ctgtcaccg tcattagtc ctcttgcgt 16980

ttgggtttat gttgattcta aaggttacag cccagttacc agtgtatgtt actgtgactg 17040
 tgtaaatcgt gtttactgcc tagcccaggg tgtgccaggg taaagttcct tctccaaagt 17100
 cccagtgtc taaccttac cccctcaaa ggagttttt ttgtttttt tttttttt 17160
 ttttgaga cagagtctca ctctgtggcc caggctggag tgcagtggca cgtctctgc 17220
 tcactgcaac ctccgttcc caggtttaag caattctcat gcctcagtct cccaggtagc 17280
 tgggaatata ggcattgcc accacgcca actaatattc tatacttta gtagagacag 17340
 ggtctacta tgttgccag gctggtctca agtgatctgc ccacctgggc ctccagtaa 17400
 tttttttt gagatagtg ttactcttg ttgccatgc tgggtgcaat 17460
 ggtgtgatct tggctaccg caacctccg ccccagggt ccagcaatta tctgcctca 17520
 gcctccaag tagctggat tacgggtgtg caccaacatg cctggctaatt ttgtattt 17580
 tagtagagat ggggttctc catgtgatc agactggtct caaactcctg acctcagga 17640
 atccgccac ctacgctcc caaagtgtg ggattacagg tgtagccac catgctcagc 17700
 cctcagtaa ttctttttt ttttttcc gagacggagt ctgctctgt cgtccaggct 17760
 ggagtgcagt ggcgtgatct cggctcactg caagctccg ctctgggtt cagccattc 17820
 tctgcctca acctcccgag tagctggac tacaggcgtc cgtccaccag cccagctaatt 17880
 ttttgatt tttagtaaaa acgggggtt caccgggggt ctgcatctcc tgacctggg 17940
 atctgccctg cctaacctc tcaaagggt gggattacag gctggagccc ctgggccccg 18000
 gccgaattt ttaatggaa aaacaaggc tcaatttggg ggccaaggct gacctcaaac 18060
 tctgagtc aaggatcct cctgcctcg gttnnnnnn nnnnnnnnn nnnnnnnnn 18120
 nnnnnnnnn nnnnnnnnn nnnntggtct gggctaatt ttgttttt gtagagagag 18180
 tattgccat gttgccagg ctggtctga actcctggg tcaagcatc ctctgcctt 18240
 ggctctaaa agtctcga ttatagctg gagccacct acctggcctg attacatgtt 18300
 actgtacct ttccactct gtgatact caagcacatg ttacacgca catatatgt 18360
 aatatttag gtgtgattc ttgatttac aaaactatt gccgtgtga gaaatcata 18420
 atctgttta cattttga atactttt agcgtgtg aatattact accaggtctt 18480
 taccgtgtt tccctgtag atctcagaat aattaccgca tcatgcca ccaggagctg 18540
 ctggccatc glaagaccc agccggga aggaagacac cagctgtgg cctccagggt 18600
 cccaggctg cccctctg tctctgcat ttaggaaga ccatgatgtt ggtgatgaac 18660
 tgggaggga aaggtggccc cagatgggt ctctggaat atttagttt gtttgggtt 18720
 ttgagatgg agtctcact tgtgcccag gcaggagtgc agtggcacga tctcgctca 18780
 ctacaacct tgctcttg gttcaagtga ttctctgctc tcagcctcc aagtagctg 18840
 gattaaagg atatgccacc atgcatggct aattttgta tttgagtag agatggggtt 18900
 tcacatgtt ggccaggct gtctgaact cctgacctca agtgalccg ccgctcagc 18960
 ctcccaaagt gctgagatta taggcgtgag ccactgcgc tggcctatgc cttgtatct 19020
 taaacctga gactcagat gtggcctgg accagcagcc tgggcactg gcattcctg 19080
 ggagcttact ggaaaggcag gatctcagg cccacccag atctccgaat caggatctg 19140
 attctcagg gcacgcccag gggattcatg ggtcagctt agaagtcaaa gttggaagc 19200
 cgggcatgtt ggtcacgc tgtaatcca gcacttggg aggcagagg aggcagatca 19260
 cctgaggtca ggagttgaa accagctgg ccaaacccg tctctactaa aaatacaaaa 19320
 attagctggg tgtgtgtgt tgtcctgt atccagcta ctgggaggt taggcagga 19380

gaatcacttg aacccgggta ggcggagggt atagttagcc aagattgcac cactgcactc 19440
 cagcctggca acagagcgag actcctctc aaaaaaaaaa aaaaaaagt caaagtttg 19500
 gaaatgcta gagaccccat gttttcaaa gacttgttt gtatgacttt atgagatgat 19560
 caagcagggt tggccacaag gggagacagc caaaggctca ggaagataac cagttttt 19620
 ttactcaaag gtcccttagt taccagggag gccacagcat gcctacttag gcccatgtca 19680
 gaggacacca ggggtgtccc cggaaggata aatccaatc tgtgattta ggctccgcc 19740
 cttttcggg aattttctt gcttaagct aattcaata aaggccccc gcttaagact 19800
 ttcatatgt tattaattt aggcgggttc taactccta atgaagaaat attcctctt 19860
 ccctggaltt ccgaatcaca ccattcttc gagaaacgcc tccacnnnn nnnnnnnnn 19920
 nnnnnnnnn nnnnnnnnn nnnnnnnnn nnnnnngtt tttttggg caccggcca 19980
 tcatctggg aggcagcatt aaggccatc acttgccct tcaatgtgac tgtcatctt 20040
 ttacacgaaa aagtttttg gtgaaaccc ggtggtcggc ccattacct cctgagtgcc 20100
 tagggttgt gttccgggcc atttatctca gtttcaccc cctctggat gtaagatgc 20160
 tcccctctg tctgctggg gaaggcgggt cctggtcca caggatagat tgggtcaaa 20220
 gtgttcaga gctaggatct ctccaggct ctggtgggc actcctgtt cccagcact 20280
 tgggagggt agccgaatgg gtcacctaag gtcaggcatt tgagaccaga ttggccagca 20340
 tggcacactc cctgtgtgt actaaaaata gaaaaattac actgggtgtg gtggcacatg 20400
 cccgtagtcc cagcacttg gcagggtcaa ccgggtggat cactgaggtc aggagttga 20460
 gaccagcctg gccaatatgg caaaactctg tctgtactaa aaatacagaa attagccagg 20520
 cgtgtagtg tgtcctgta atcccagcta ctcaggaggc tgaggcagga gaatcgctg 20580
 aactgggagg cggaggtgc agtgagctga gatcgcgcca ctgcattcca gcctgggcaa 20640
 caaagccaga ctctgtca aaaaaaaaaa aaaaaaaaaa aaaggctagg cttacgtct 20700
 gcaagaatgt ggctgttt ctctctccc tgaaggagt cgtaggcca ggccccagg 20760
 atgattccc cgagccctgc ctctcctt cctctggg tttgtccc cctcccatc 20820
 tcttctct ccactctg tgcttccct ccacgatcag cctgtctgc ctctcccca 20880
 ggtctacca acatgttggg ctccctgc tctctacc cgtcacagg cagcttga 20940
 cggtagtga catgtccgc tctctgtt gccacgtg gacgcctaa cgtgttacg 21000
 ctgacaagga gtctcctgc cctgacccg ggcggcgtc ctccactgt aacgtccgt 21060
 ggagaggcag ggctgggggt caccactgt cctccagggt gtctctgt tcttattct 21120
 catagatct cctgcagtt catactagaa agttccact ggcatgttg tacaccctg 21180
 ttatccatt tactctggag gcggaagcag gagaatcgt tgagccagc aggatgaggc 21240
 tgcagttagt tgtaccgca ccgtgcact ccagcctgg caacagagcg agaccctgc 21300
 tctaaataa atactagaaa gtcccagca cgcaaagcc ctctagctc ctggtgccag 21360
 agtcaattcc tgaaaggacg tggagatagg aagggcctcg gctctgct gaagcagcg 21420
 ggcatgaagc ttgcccaga tgcctacgg cccctcctca gtcaggacaa caggatggag 21480
 gtgacctgt gctaaagga gagaaggagg cgtcgcttg cactgccag tccccagct 21540
 ggtgacctt gccctgctg gtatggggc cccacctga tggggcagg agacagatc 21600
 ggcaggaacc tgaaaggaca cgtgctct gagctctc ctatagtcag ggtggccaa 21660
 gcgcggtgt ctgtactgc accctaagtc tcttgctc ggtccctg cagtcccg 21720
 ctgctcca agcgtgct ggagctgacg tccctcga agatcagca caggagtgt 21780

gactgaggtc tccctttcc cggccccctgg tgactgacgg tctctgtgt gcctccagg 21840
 acagccgtga acgctcagtc ggggggtgac accccggcgg ggggcctggt gacgggtaag 21900
 gcccccatc tccccttgt gccgcagcc ctgagagtgg gagaaagga ggagggggcc 21960
 cacagagacg tcccttggc tcatgggccc tgcgccccgg gactgcacag ggacttggg 22020
 ggccacacag gagtaggggg accacaggag actgagcagg ggcgtggggc ctggcagtc 22080
 gtgccttac cccaccctt glccccagt ggctctgtg aacaagaggc tgctacgtg 22140
 cgtgctggg ggacctgca ctccgaggt cactgtgtt cccgtgccc gcaggagtgc 22200
 ttgtctgtct gtcctggac tacctgacct cactgttcta ctacatccc aagtctgcc 22260
 tggctgccgt catcatcatg gccgtggccc cgctgttga caccaagatc ttcaggacgc 22320
 tctggcgtgt taagagtacg tcctgtctt acaggggaga gcgctgtgat gcggtgtctg 22380
 aacggaggag gtgtcattta tgctaccca tttcctgca gccccctgt tggggctggg 22440
 actgggaagt tagggcagtc ccggaacaga gaagtggatg gccaggagat ggccccagag 22500
 atgttccga ggctcagtg gaagagctgg agctcctgt cctgacacct ggggtctga 22560
 ggcgagcact gacccggggg agggccccct cctgatccc ctgccccat cctaccctc 22620
 ctggccccc gctccagcc accactctgc ccggcccagc tggggggagg gacaggagac 22680
 gtccctggg accagcagg ccagcggaa acgcttgac cctggctcag aatggcagtt 22740
 cctttttt ttattattat tttttatt tttttatt gatcattct ggggtttct 22800
 tgcagagggg gatttggcag ggtcatagga caatatggg ttgggggtaa ggtcacagat 22860
 aacaggatcc caaggcagag gaattttct tagtcagaa caaatgaaa agtctccat 22920
 gtctactct ttctacacag acacagcaac catccgatt ctcaatctt tccccactt 22980
 tccccctt ctattccaca aaaccgcat tgcacatg gcgcttctc agtgagctgt 23040
 tgggtacacc tccagacgg ggtggggcc gggcagagg gctcctcact tccagtagg 23100
 ggcggccggg cagaggcgcc cctcacctcc cggacgggag cagttcctt aactaacac 23160
 atttggttt gttgtggaa gaagcatgga gtgggttac ctacttact tggcatttc 23220
 tggtagccc cagcaagtg ccaggtagaa ggaatgataa tttctgca tggcgtgc 23280
 atgccagggc ctttcgctg ccagtgcgt gggccttag aacagccctg aaacgtgtc 23340
 atgtccggc gtggaatgg atggccgctg ctgctagaaa ccaaggctca gccagggtg 23400
 ggcaggcctg acctatcgt gcggtggaat cccccacagg gctaagcccg tgcactttgt 23460
 cccaggga ccttctctt ggccaggct caagggtca cgtgttccct gcccactcc 23520
 tcaggccagc tctgtccct gacaagcccc tgctgtgcc ctccctgagg ttggaggcca 23580
 gggaccggc gccaggtct accgtgctc agctcagatg gggagggcat tctttctt 23640
 cgactgaag catggcctg tcagcagctg ctgtcccaa gtctcagg gctgcttgg 23700
 gtccatgag acctttac atagtggg ggcagaaggc tgtccgctg ttcagcagg 23760
 ccatgtggg gctcgggca gctccgggc attcctcagc tgtccctc tctagggt 23820
 ggacctgtg cccctgtcg tgacctct gctgtctt tggagggtc agtacggcat 23880
 cctggccggg gccctgggt ctctgtcat gctctgac tctgcagca ggcctgagac 23940
 caaggtaccc ctccgtggc tctagtggt gagtgtctg ggggcaggat tctgggcat 24000
 ggtctatgt ttgagggc cgggtgatt gtgtcgtg gtgtgtga aggggaccg 24060
 tcgtcggcag tggggcagc acctgtct aaacctggt gttctccac tgttgggg 24120
 ccgtgggggt ctcccttag caccctctc ccggcccc gcagcacgt aggtgggtg 24180

ggggctccc gcttgggaca ggccaagcct ggtggaggcc acccggtcag acccgctcc 24240
 aggactcact cctccccaca ggtgtcagag gggccgggtc tggctctgca gccggccagc 24300
 ggctgtcct tccctgccat ggaggctctg cgggaggaga tccaaagccg ggccctggaa 24360
 ggtgcatggg cgggggtcaa ggtggtctga ggtcactccc ctgtctctg cccccactc 24420
 cctgtcttc aggaccccaa gacctgtcc ccgacgtct ccagtcaca aggatgcagg 24480
 catctctgag tgggtggac cgtctctgt gggcctcagc cagtggctgc tgcagcaagg 24540
 gtgtggctc cccacatac actcctccc tggccctaaa gtccggtcc tgttctggg 24600
 gggttgatt taggggagct aagggcctgt ggtcctagg agggaaacag ctgtctctgt 24660
 caccaaaca ttgtctctg tctgccacc cgaatcccc aactgggca ctacagccg 24720
 acgagatgga gcactctggc ctctctgcc cctgccctg gccagagcct ctttggcct 24780
 ctgcagagca gcttgggct gctctgggtg gctgacctg gctgggctc gctcccccag 24840
 tctccccgc acgtgcctg tctctggagt gcacccatgt ctgcagcatc gactacactg 24900
 tgggtctggg actcggcgag ctctccagg actccagaa gcagggcgtc gccctggcct 24960
 ttgtggcct gcagggtggg gtgcactggg atgccttagg ggtagcagc tgcggaagg 25020
 ccttctgtg cctgcctccc atggcgaatg tgacatctt gggctgtgat gctggacggc 25080
 ccttcggcg gtgtggctc tcttctgat ttaaacagtt ctgtcccca tctggcctc 25140
 ctgtccctc cctgtggaag ggggagcggg ggccccagc cctccgaggg gtcacgttat 25200
 ggcttctgt cactgccaca ctgtccttg tctggggac acacagtga cgagggtcag 25260
 tccctgtct cagggaatt gtatttagg aaacaaaaag agacaggtga ggctaggcgt 25320
 ggtggcacac tcttgaatc ccagcactt gggaggcaa ggcgggcaga tcacctgagg 25380
 tcaggagttc aagaccagc tggccaacat ggtgaaacc cgtctctact aaaaatacaa 25440
 aaaaaactag tcagggttg tggcgagcg ctgtaattc agctacatg gaggttggga 25500
 gaatcgctt aacctgggag gtggaggtg cagtaagcc agatccact actgtactc 25560
 agcctggcg acagagtggg actccatctc aaaaacaaac aaacgaaaa aaacaggtgc 25620
 tcatagaatt tcatgaaaa cgtatttca gggctccag aggtgaaga cgggttcta 25680
 tggaggcgt cctgtcaga gccgcagga aagtgaagg gctgggtccc aggccctgcg 25740
 tcttagcct cacctaggag cctctgagc actgcagggt caacatcca ggggtgtggc 25800
 cagtgtttc agagaggcag ggtctctgt tctgtggtt aatgtgcgc tcttgccaa 25860
 gtctagatg gcagaacgtg ggggactagg gctgtcccc aggcgcccag aggagacatt 25920
 catgaactag ccatggaaca ggaggccaag tgaccctat accccagggtg tggaccacag 25980
 ccgaccctg tcagagttc ctctctccc aacctggccc acgaggctag tttatcttc 26040
 atccccatg ccagaagagg tcaactaggc ccagagaagc caggtaatct gccacggtc 26100
 acacaggatg gtgggccag ctctcagatt gggcatgtgc ctccagaact tgcctcataa 26160
 gtgtacgtc ctgtgcata ttgcagagc acatgtctc atggtcttg gagatgattc 26220
 acttaaaaa atactctct gaccagaca tggtggtca cacctgtaat ccagcactt 26280
 tgagaggccg aggcgggagg tcaagggtt gagaccagc tggccaacat agtgaactc 26340
 cgtctctact aaaaatacaa aaattagac ggcatgttg cgcacgctg taatcccagc 26400
 tgcacgggg gctgaggcag gagaatcgt tgaaccggg aggcagaggt tgcagtgcg 26460
 cgagatcgc ccattgcact ccagcctggg taacaagagt gaaactccat ctcaaaaaa 26520
 aacaaacaaa caacaaaaa gtcctctgg ccaggcgctg tggctcacac ctgtaatccc 26580

agcactttgg gaggccgagg taggaggatc attgaggcc aggagttga gaccagcctg 26640
 ggcaacatga cgaaccccca tcaaaaaaat aggaaaaaat agctggacgt ggtggtgtgc 26700
 acctgtagct actcaggagg ctgagggtgg aggatcactt gactctggga ggttgaggct 26760
 gcagttagcc atgatcgtc cactgcactc cagcctgagc aatgagcaag accctgtctc 26820
 aaaaaacaaa aaatttaaaa aaaaaaaatc ttctctgac agcattcccc tggggctgcg 26880
 ttctctca ccattcactg gtatggagggt gaagccatac ctctccggga gactctgaga 26940
 .tggcatgtct accaggctgc cgaccgtgt gctacagagg aacatccctg ccctggctaa 27000
 agtctgtctg tctctcagggt ccccgctctc cgtgtcctgc tgtccgtga cctgaagggg 27060
 ttccagtact tctctaccct ggaagaagca ggtgggcaca gtcagacatc ctgtggcttt 27120
 ggtgattttg taaaaatcat aaatgcttat tgaaaaaat atgggaaacc agctgggcac 27180
 agtggctcat gcctgtaac ccagcacttt gggaggccga ggcagggtga tcacctgagg 27240
 tcaggagttc gagaccagcc tggccaacat ggtaaacct catcagtact aaaaatacaa 27300
 aaattagctg ggtgtggtg catgcacctc tagtcccagc tattcgggag gctgaagcag 27360
 cagaatcgct tgaaccagg aggagagggt tgcactgagc cgagattgtg cctctgact 27420
 tccagagtag ggtgagagag ctagccatgt tgtccagaaa annnnnnnnn nnnnnnnnn 27480
 nnnnnnnnn nnnnnnnnn nnnnnnnnn ncttgggtga gcaggataa tgccttgcaa 27540
 cccaggtagc agcagttgta ctgagtcga gaattgtcc actgcactcc agcctgggtg 27600
 agagcagcga gtactctgc tcaaaaaaaa aaaaaacaaa aaaaacagga aaccacctcc 27660
 accccagtc actttgtga tcaactcata cccctccct aaacacgcat atgtacctgc 27720
 ctgtcaggat gtgatatat gttgcggt ttacgtaat gggaccattt calacctgt 27780
 gctclggaac ccacatttt catcagaag gtggaagga tgccttcca gccgaaagtc 27840
 cacatcccg gggatcagga cagagcagg cgggtcagg agatccagaa gcccgggac 27900
 agaaggtagc ggaggacag gagcagggt ggcgtgacc ctgagacag caatcgagg 27960
 aggtctgagc cgcagcagg gtcaacaaga ggaaggcca gagatgcaga gcacccacc 28020
 caggccacac agcagtgcc agagggtccc agggcccagt gctaggcctc ttctcttcc 28080
 actgaggta cagctgaagc tgggtcagct ccgtgagagt gagggtggc ggaattgta 28140
 ctgaattct tggctcaatg tgacgtcagg gagatcacc catgtgttg tagaatcagc 28200
 tcagatgcag tgcactggag gtgataagc agaattgtgc ctggcgtg gatgggagg 28260
 taccctggag cagtaagaaa gggccgttag tcacctgaaa aatacgctta cagagactca 28320
 ggtgagacc tcattggagt agtgacatg gcctgggtg cccacagctc ctctctcac 28380
 acctccagg actctggaag gccctctta atccctct gtgaactgac ccacctcac 28440
 ttctgagct ttagtgctg aaacattat tttatttct gcagagaagc acctgaggca 28500
 ggagccaggg acccagccct acaacatcag agaagactcc attctggacc aaaagggtgc 28560
 cctgtcaag gcataatgg gccacccgtg ggcacccaca gtttcagggt tgtccgga 28620
 ggttctgtc actgtgatt gatgtggat gccgctgat agacatgtc gcctggctga 28680
 gaaacccctg agcaggaac ccagggaaga gaaggagcc aggcctggag gtccacggca 28740
 gtggggagt gggctcactg gcttctgtg ggaatgactg aaaatgacct cgtgtctgt 28800
 ccctggcatg accctcttg gaagagtgt ttgagagag cctctagaa tgacagactg 28860
 tgcaggaag caggggcagg ggttccagc ccgggctgt cgaggcatcc tggggctgc 28920
 agcaccttc cggctacca gtgccactg cgggggagg acggggcagg caggagtctg 28980

ggaggcgggt ccgctcctct tgtctgccc atctgtgctc tccgagagaa aaccaagggt 29040
tgtcaaatga cgtcaagtct ctatttaaaa ataattttgt gttttctaaa tggaaaaagt 29100
gatagctttg gtgattttgt aaaagtcata aatgcttatt gtaaaaaata caggaaacca 29160
ccccctcccc tgtccacttg ggtgatcatt ccagaccctt ccccaaaccat gcatatgtac 29220
ctgtccgtca gtgtgtggat gtatgtttac agttctacat aaatgggac atttataca 29280
tgggtgctcg gaaccacat tttcatgca gtcatgtgca gtgaattatt tattgtgata 29340
ataaatagca ttagaataca agattttta tgtctgctg gtattttgt ctatatatgc 29400
atcataattg acttaccgag cctctgttc aacgtgtgctg tgggttagag acgggatcgt 29460
gcctccttta gatgtgtcag tgaagccct tgcctctat gtgactgctg ttgaaacag 29520
ggcttttagg gggtaactag gtttaataa ggtcataaag gtggccctg aaccgatagc 29580
tttagtgtcc tcgtgagaag agacgccaga gagctcgtc tctgaccct cacaccccag 29640
ggaactccat gagaggacac ggcaaaacca ggccatgcgc ccaccaggaa gaaaggcctc 29700
atgaggaccc cagcctcaa aactgggaga agatgaatct ctgtgctagg cccgcagcc 29760
tggggtgatc tgtacggcg gcctgagcag gtgaggactg cctgcatgtt tgtattttta 29820
tgaatgcttt gattgagctt gggggtaa at cctggaggc ctgtggcagc ctacagggtg 29880
tgtctccct gcacttctt caagaagaat gtggcctgcc ctgtgagcc tcgtctgcc 29940
cgttctgcc gggcagtc ccgaatgtc agcgcagcaa ggggagggcc tctgtaacca 30000
gggctgctgg ctgctgggct cccactgga caggggagcg gacattggag tgccttcat 30060
ccgctgact ctccaggtc cctgcctcc catcttctt ccttcttcc tctttctt 30120
cttcttctt tcttcttct gctccttcc ttttcttct tcttctggt ttttcttct 30180
ttcttctt tcttcttct tcttcttcc tcttcttct tcttcttct ccacccctt 30240
cttctcactg tgttccccag gctgttctt aactcctggg ctcaagcaat cctcttgcct 30300
cagcctctgg tgtggctgg accacagaca catgccaccg cgccaggcta attgttaat 30360
ttttttata gagacagggt ctacttgtt tcccagggt cgtctcaaac tctgggctc 30420
aagtgtact ccagctcag cctgaagtgc tgaattaag gtgtgagcca ccatgcctgg 30480
ccctcttca ttcacaagt aaccgttcac cctgcctc caatccatgt cgttctgac 30540
ctcaggcagc tttcttct acataaagt aacctgtccc aaagctgtgc tactgggcc 30600
ccccgccag ggttgaaga ggcagcagtt cacattggc ttgactcac ataccaagg 30660
catccatgct tgaacctaga catggttcat tcacagggt ggtgaagg aaacagactg 30720
tggcgggcac taggactat cagttcata aatctgata ccactgtga ctacagcgtt 30780
atgcttctg gaatctact aaacatgct gagtcagct tcaggaagc aggcacaagg 30840
acgtttgtga caactggc agtaattta agaagtagga agaaacctga gtttctga 30900
attcggtaac atctgactg taggacacac gaataatgcc gtggaatatt gtgcagcagt 30960
taaaagaaat gagggaagg cgggtgtgg cagctcacgc ctgtatctc agcacttgg 31020
gaggccgagg ctgttgatc acttaaggc aggagttga gaccagtctg gccaacatgg 31080
cgaaacctg tcttactaa aaalacaaaa attagctgac ctctgtgaca gagtgagact 31140
ctgcctcaaa aaaaaaaaaa aaaaaaaatgagg tgaagctta agttgtaaca 31200
ctgatttgg ggtgcaata agcaagtgc agaataatcc ccatgttaag atgctattg 31260
agtgaacaca cggaccaaac aattctatgt tgggtacaaa ttaagaaag agttctggc 31320
tgggcacggg ggtcacacac tgtaatcca gcactctgg agggtagtc ggatgatca 31380

ggaggtcagg acttctagac tagcctggcc aatatggtga aaccccatct ctactaaaaa 31440
 taaaaaatt agccaggcgt gttggcgcgt gtcggagtc ccagctactc aggagactaa 31500
 ggtaggagaa tcacttgaac ccgggaggca gaggttcag tgagctgaga tcatgccact 31560
 gcactctggc ctgggcaaca gagcgagact caaaaaataa ataaataaat aaataaaata 31620
 aataagagtt ttctgcactt tgggaggcct gtagtcccag ctactctgga ggctgaggca 31680
 agaggatcac ttgagcctgg gggggtcgag gctgcagtga gtctgattg tgcactgaa 31740
 atccagcctg ggcaacagag tgagac 31766

<210> 6

<211> 548

<212> PRT

<213> Human

<400> 6

Cys Arg Pro Ser Thr Val Thr Asn Lys Phe Pro Ile Leu Lys Trp Leu
 1 5 10 15
 Pro Arg Tyr Arg Leu Glu Tyr Ile Met Gln Asp Phe Ile Ala Gly Phe
 20 25 30
 Thr Val Gly Leu Thr Thr Ile Pro Gln Ala Ile Ala Tyr Gly Val Val
 35 40 45
 Ala Gly Leu Glu Pro Gln Tyr Gly Leu Tyr Ser Ala Phe Met Gly Cys
 50 55 60
 Phe Thr Tyr Ile Val Phe Gly Ser Cys Lys Asp Val Thr Ile Ala Thr
 65 70 75 80
 Thr Ala Ile Met Ala Leu Met Val Asn Gln Tyr Ala Thr Ile Ser Pro
 85 90 95
 Asp Tyr Ala Val Leu Val Cys Phe Leu Ala Gly Cys Ile Val Leu Leu
 100 105 110
 Leu Gly Leu Leu Asn Met Gly Val Leu Val Arg Phe Ile Ser Ile Pro
 115 120 125
 Val Ile Thr Gly Phe Thr Met Ala Ala Ala Thr Thr Ile Gly Ser Ala
 130 135 140
 Gln Ile Asn Asn Ile Val Gly Leu Thr Ser Pro Ser Asn Asp Leu Leu
 145 150 155 160
 Pro Ala Trp Lys Asn Phe Phe Thr His Leu Thr Ser Ile Arg Leu Trp
 165 170 175
 Asp Ala Leu Leu Gly Val Ser Ser Leu Val Phe Leu Leu Leu Met Thr
 180 185 190
 Arg Val Lys Asp Ile Lys Trp Gly Asn Arg Ile Phe Trp Lys Tyr Leu
 195 200 205

Gly Leu Ser Arg Asn Ala Leu Ala Val Ile Phe Gly Thr Phe Leu Ala
 210 215 220
 Tyr Ile Leu Ser Arg Asp Gly Asn Gln Pro Phe Arg Val Thr Gly Asn
 225 230 235 240
 Ile Thr Ala Gly Val Pro Pro Phe Arg Leu Pro Pro Phe Ser Thr Thr
 245 250 255
 Val Asp Gly Glu Tyr Val Ser Phe Gly Glu Met Ile Ser Thr Val Gly
 260 265 270
 Ala Ser Leu Gly Ser Ile Pro Leu Ile Ser Ile Leu Glu Ile Val Ala
 275 280 285
 Ile Ser Lys Ala Phe Ser Lys Gly Lys Ile Val Asp Ala Ser Gln Glu
 290 295 300
 Met Val Ala Leu Gly Met Cys Asn Ile Met Gly Ser Phe Val Leu Ser
 305 310 315 320
 Met Pro Val Thr Gly Ser Phe Thr Arg Thr Ala Val Asn Asn Ala Ser
 325 330 335
 Gly Val Lys Thr Pro Leu Gly Gly Ala Val Thr Gly Ala Leu Val Leu
 340 345 350
 Met Ala Leu Ala Phe Leu Thr Gln Thr Phe Tyr Phe Ile Pro Lys Cys
 355 360 365
 Thr Leu Ala Ala Ile Ile Ile Ala Ala Met Ile Ser Leu Val Glu Leu
 370 375 380
 His Lys Ile Lys Asp Met Trp Lys Ser Lys Lys Lys Asp Leu Phe Pro
 385 390 395 400
 Phe Val Val Thr Val Leu Thr Cys Met Phe Trp Ser Leu Glu Tyr Gly
 405 410 415
 Ile Leu Cys Gly Ile Gly Ala Asn Met Val Tyr Ile Leu Tyr Ser Ser
 420 425 430
 Ala Arg Pro His Val Asp Ile Lys Leu Glu Lys Ile Asn Gly His Glu
 435 440 445
 Val Ser Val Val Asp Val Lys Gln Lys Leu Asp Tyr Ala Ser Ala Glu
 450 455 460
 Tyr Leu Lys Glu Lys Val Val Arg Phe Leu Asn Asn Gln Asn Gly Glu
 465 470 475 480
 Thr Gln Leu Val Val Ile Lys Gly Glu Glu Ile Asn Ser Ile Asp Tyr
 485 490 495
 Thr Val Ala Met Asn Ile Val Ser Met Lys Gly Asp Leu Glu Ala Leu
 500 505 510
 Asn Cys Ala Met Ile Cys Trp Asn Trp Asn Ile Ala Ser Ala Gly Val
 515 520 525

Val Cys Arg Leu Asn Asn Asp Leu Arg Pro Ile Phe Lys Phe Asp Leu .
 530 535 540

Ser Leu Glu Glu
 545

<210> 7

<211> 548

<212> PRT

<213> Drosophila melanogaster

<400> 7

Cys Arg Pro Ser Thr Val Thr Asn Lys Phe Pro Ile Leu Lys Trp Leu
 1 5 10 15
 Pro Arg Tyr Arg Leu Glu Tyr Ile Met Gln Asp Phe Ile Ala Gly Phe
 20 25 30
 Thr Val Gly Leu Thr Thr Ile Pro Gln Ala Ile Ala Tyr Gly Val Val
 35 40 45
 Ala Gly Leu Glu Pro Gln Tyr Gly Leu Tyr Ser Ala Phe Met Gly Cys
 50 55 60
 Phe Thr Tyr Ile Val Phe Gly Ser Cys Lys Asp Val Thr Ile Ala Thr
 65 70 75 80
 Thr Ala Ile Met Ala Leu Met Val Asn Gln Tyr Ala Thr Ile Ser Pro
 85 90 95
 Asp Tyr Ala Val Leu Val Cys Phe Leu Ala Gly Cys Ile Val Leu Leu
 100 105 110
 Leu Gly Leu Leu Asn Met Gly Val Leu Val Arg Phe Ile Ser Ile Pro
 115 120 125
 Val Ile Thr Gly Phe Thr Met Ala Ala Ala Thr Thr Ile Gly Ser Ala
 130 135 140
 Gln Ile Asn Asn Ile Val Gly Leu Thr Ser Pro Ser Asn Asp Leu Leu
 145 150 155 160
 Pro Ala Trp Lys Asn Phe Phe Thr His Leu Thr Ser Ile Arg Leu Trp
 165 170 175
 Asp Ala Leu Leu Gly Val Ser Ser Leu Val Phe Leu Leu Leu Met Thr
 180 185 190
 Arg Val Lys Asp Ile Lys Trp Gly Asn Arg Ile Phe Trp Lys Tyr Leu
 195 200 205
 Gly Leu Ser Arg Asn Ala Leu Ala Val Ile Phe Gly Thr Phe Leu Ala
 210 215 220

Tyr Ile Leu Ser Arg Asp Gly Asn Gln Pro Phe Arg Val Thr Gly Asn
 225 230 235 240
 Ile Thr Ala Gly Val Pro Pro Phe Arg Leu Pro Pro Phe Ser Thr Thr
 245 250 255
 Val Asp Gly Glu Tyr Val Ser Phe Gly Glu Met Ile Ser Thr Val Gly
 260 265 270
 Ala Ser Leu Gly Ser Ile Pro Leu Ile Ser Ile Leu Glu Ile Val Ala
 275 280 285
 Ile Ser Lys Ala Phe Ser Lys Gly Lys Ile Val Asp Ala Ser Gln Glu
 290 295 300
 Met Val Ala Leu Gly Met Cys Asn Ile Met Gly Ser Phe Val Leu Ser
 305 310 315 320
 Met Pro Val Thr Gly Ser Phe Thr Arg Thr Ala Val Asn Asn Ala Ser
 325 330 335
 Gly Val Lys Thr Pro Leu Gly Gly Ala Val Thr Gly Ala Leu Val Leu
 340 345 350
 Met Ala Leu Ala Phe Leu Thr Gln Thr Phe Tyr Phe Ile Pro Lys Cys
 355 360 365
 Thr Leu Ala Ala Ile Ile Ile Ala Ala Met Ile Ser Leu Val Glu Leu
 370 375 380
 His Lys Ile Lys Asp Met Trp Lys Ser Lys Lys Lys Asp Leu Phe Pro
 385 390 395 400
 Phe Val Val Thr Val Leu Thr Cys Met Phe Trp Ser Leu Glu Tyr Gly
 405 410 415
 Ile Leu Cys Gly Ile Gly Ala Asn Met Val Tyr Ile Leu Tyr Ser Ser
 420 425 430
 Ala Arg Pro His Val Asp Ile Lys Leu Glu Lys Ile Asn Gly His Glu
 435 440 445
 Val Ser Val Val Asp Val Lys Gln Lys Leu Asp Tyr Ala Ser Ala Glu
 450 455 460
 Tyr Leu Lys Glu Lys Val Val Arg Phe Leu Asn Asn Gln Asn Gly Glu
 465 470 475 480
 Thr Gln Leu Val Val Ile Lys Gly Glu Glu Ile Asn Ser Ile Asp Tyr
 485 490 495
 Thr Val Ala Met Asn Ile Val Ser Met Lys Gly Asp Leu Glu Ala Leu
 500 505 510
 Asn Cys Ala Met Ile Cys Trp Asn Trp Asn Ile Ala Ser Ala Gly Val
 515 520 525
 Val Cys Arg Leu Asn Asn Asp Leu Arg Pro Ile Phe Lys Phe Asp Leu
 530 535 540

WO 02/059306

PCT/US01/42809

Ser Leu Glu Glu
545